

ROWENA MARIA TEIXEIRA VIEIRA

**IDENTIFICAÇÃO DE INDIVÍDUOS PELA DINÂMICA DO
CAMINHAR**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. Evandro Ottoni Teatini Salles.

Co-orientador: Prof. Dr. João Marques Salomão.

VITÓRIA
2009

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

V658i Vieira, Rowena Maria Teixeira, 1984-
Identificação de indivíduos pela dinâmica do caminhar /
Rowena Maria Teixeira Vieira. – 2009.
79 f. : il.

Orientador: Evandro Ottoni Teatini Salles.

Co-Orientador: João Marques Salomão.

Dissertação (mestrado) – Universidade Federal do Espírito
Santo, Centro Tecnológico.

1. Sistemas de reconhecimento de padrões. 2. Markov,
Processos de. 3. Análise de componentes principais. 4. Análise
de variância. 5. Locomoção humana. I. Salles, Evandro Ottoni
Teatini. II. Salomão, João Marques, 1956-. III. Universidade
Federal do Espírito Santo. Centro Tecnológico. IV. Título.

CDU: 621.3

ROWENA MARIA TEIXEIRA VIEIRA

**IDENTIFICAÇÃO DE INDIVÍDUOS PELA DINÂMICA DO
CAMINHAR**

Dissertação submetida ao programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para a obtenção do Grau de Mestre em Engenharia Elétrica.

Aprovada em 19 de dezembro de 2009.

COMISSÃO EXAMINADORA

Prof. Dr. Evandro Ottoni Teatini Salles - Orientador
Universidade Federal do Espírito Santo

Prof. Dr. João Marques Salomão - Co-orientador
Instituto Federal do Espírito Santo

Prof. Dr. André Ferreira
Instituto Federal do Espírito Santo

Prof. Dr. Rodrigo Varejão Andreão
Instituto Federal do Espírito Santo / PPGEE – UFES

A diferença entre o possível e o impossível está na vontade humana.
(Louis Pasteur)

Dedicatória

À minha mãe Maria Luiza pelo amor e carinho, e por ter me proporcionado uma base educacional digna, pois foi a partir desta base que eu busquei cada vez mais avançar e conhecer novos mundos. E, sobretudo por ser um exemplo de luta e perseverança.

Ao meu pai José Orlando, que mesmo distante sempre torceu pelo meu sucesso.

Aos meus demais familiares por estarem sempre presentes onde quer que eu esteja.

Ao meu namorado Francis, pelo apoio e incentivo para que eu sempre continuasse em busca de meus objetivos, e pelo afeto que me faz sempre querer ser uma pessoa melhor.

Agradecimentos

Quero agradecer, primeiramente, à Deus, por me conceder o direito à vida e por me proporcionar a busca pelo conhecimento, a fim de eu possa compreender melhor o mundo ao meu redor.

Agradeço à CAPES, e ao Programa de Pós-Graduação em Engenharia Elétrica da UFES e membros participantes pelo financiamento concedido e pela confiança em mim depositada para desenvolver esta pesquisa.

Em especial ao meu orientador e professor Dr. Evandro Ottoni Teatini Salles, pelos ensinamentos e pelos valiosos conhecimentos passados, pela paciência e pela solidariedade nos momentos mais difíceis e decisivos ao longo destes últimos anos. Ao meu colega de trabalho, co-orientador e professor Dr. João Marques Salomão pela ajuda e direcionamento no desenvolvimento deste trabalho, e por me permitir dar continuidade ao seu trabalho iniciado na sua tese de doutorado.

Agradeço também aos meus amigos do Laboratório CISNE, LEPAC e LAI da UFES, brasileiros e peruanos, pelas trocas de informações, pela preocupação, pelos momentos de estudos em conjunto e também pelos momentos de distração em que nos reuníamos para discutir assuntos do nosso cotidiano e de nossas vidas pessoais.

Finalmente, agradeço a todos aqueles que torceram para que este projeto se tornasse realidade!

Resumo

Os sistemas de segurança baseados em identificação de indivíduos através de características biométricas têm exigido que os programas de computador responsáveis por esta função sejam os mais confiáveis possíveis. As características biométricas como face, íris e impressão digital necessitam de recursos sofisticados e da ajuda direta do indivíduo que se deseja reconhecer. Baseado neste contexto, buscou-se uma maneira de minimizar estas exigências através da característica biométrica do caminhar.

O caminhar humano é característico de cada ser humano, e este aspecto pode ser utilizado de forma favorável nos sistemas de reconhecimento. Assim, neste trabalho foi desenvolvido um sistema de reconhecimento de indivíduos pela dinâmica do caminhar, usando a abordagem holística para a composição dos vetores de características. Para tal foi usado um estimador de fundo robusto, o LMedS, para se fazer a extração das silhuetas. Após esta etapa, foram feitas algumas filtrações nas imagens, a fim de melhorar a qualidade das silhuetas capturadas, e a medida de largura destas silhuetas serviu como característica de cada classe de pessoas.

Métodos de redução de dimensão como o PCA, LDA e POV foram testadas para que os vetores característicos ficassem mais representativos e para que o tempo de processamento do sistema fosse reduzido na etapa posterior. Como classificador foi utilizado o HMM, pois se pode relacionar os estados de um ciclo de caminhar aos estados presentes no HMM. Neste caso, como cada pessoa tem uma forma de caminhar específica (por exemplo, uns andam mais rápido, outros andam mais devagar), este classificador se enquadrava muito bem à proposta da pesquisa. Os resultados obtidos mostraram que a técnica da LDA é a mais satisfatória, pois foi a ferramenta de auxílio ao HMM que obteve taxas de acerto mais expressivas. Tendo em vista as tendências atuais de fusão biométrica, a forma de caminhar se mostra muito atrativa para ser utilizada em sistemas bi-modais ou mesmo isoladamente.

Palavras-chave: Reconhecimento pela Forma de Caminhar, Modelos Ocultos de Markov, Análise de Componentes Principais, Análise de Discriminantes Lineares, Proporção por Variâncias.

Abstract

The security systems based on identification of individuals through biometric characteristics have required that the programs responsible for this function are as reliable as possible. Biometrics modalities such as face, iris and fingerprint require sophisticated resources and the straight help of the individual which wish to recognize. Based on this context, it was sought a way to minimize these requirements through the biometric characteristics of walking.

The human walking is a characteristic of each person, and this aspect can be used favorably in recognition systems. In this work was developed a system for identification of individuals by the dynamics of gait, using a holistic approach to compose the vectors of characteristics. For this it was used a robust background estimator, the LMedS, to do the extraction of silhouettes. After this stage, it was done some filtering on the images in order to improve the quality of silhouettes captured, and the measure of the width of these silhouettes were used as characteristic of each class of people.

Size reduction methods as the PCA, LDA and POV were tested so that the feature vectors became more representative and that the system processing time would be reduced at the later stage. As a classifier, it was used the HMM, because it is possible to relate the states of a gait cycle to the states present in HMM. In this case, as each person has a specific way of walking (for example, some walk faster, others walk more slowly), this classifier fitted very well to the research proposal. The results showed that the LDA feature extraction is the most satisfactory because giving, with the HMM classifier, the most expressive hit rates. Given the current trends in biometric fusion, gait shows to be very attractive for use in bi-modal systems or even alone.

Keywords: Gait Recognition, Hidden Markov Models, Principal Components Analysis, Linear Discriminant Analysis, Proportion of Variance.

Sumário

Lista de Figuras	xii
Lista de Tabelas	xiv
Capítulo 1	
Introdução.....	15
1.1 Motivação	15
1.2 Trabalhos Relacionados	17
1.3 Objetivos da Pesquisa	19
1.4 Organização do Texto	20
Capítulo 2	
Extração do Vetor de Características.....	22
2.1 Introdução	22
2.2 A Forma de Caminhar.....	23
2.3 Extração dos Dados da Sequência de Vídeo.....	25
Capítulo 3	
Técnicas de Redução de Dimensão.....	33
3.1 Introdução	33
3.2 Análise de Componentes Principais	34
3.3 Análise de Discriminantes Lineares	36
3.4 Proporção por Variâncias.....	39
Capítulo 4	
Modelos Ocultos de Markov	41
4.1 Introdução	41

4.2 Modelos Ocultos de Markov	42
4.3 Os Três Problemas Básicos do HMM.....	44
4.3.1 Solução do Problema 1	45
4.3.2 Solução do Problema 2	47
4.3.3 Solução do Problema 3	49
4.4 O HMM Aplicado à Dinâmica do Caminhar	51
 Capítulo 5	
Procedimentos Experimentais e Resultados	53
5.1 Metodologia e Dados Utilizados.....	53
5.2 Método de Redução Dimensional Usando PCA.....	55
5.3 Método de Redução Dimensional Usando LDA	59
5.4 Método de Redução Dimensional Usando POV	66
5.5 Considerações Finais	69
 Capítulo 6	
Conclusões e Projetos Futuros	71
6.1 Contribuições da Pesquisa	72
6.2 Sugestões para Trabalhos Futuros	73
Referências Bibliográficas	75

Lista de Figuras

Figura 2.1: Diagrama de um sistema de identificação de indivíduos.....	22
Figura 2.2: Características extraídas de silhuetas binárias a partir da abordagem por modelos. (a) Medida de distâncias, (b) elipses, (c) angulação da perna, (d) combinação de formas geométricas (Nikolaos et al., 2005).....	24
Figura 2.3: Características extraídas de silhuetas binárias a partir da abordagem holística, em relação ao tempo: (a) largura da silhueta, (b) projeção vertical e horizontal, (c) representação angular (Nikolaos et al., 2005).	25
Figura 2.4: (a) Quadro de uma sequência de caminhar. (b) Fundo estimado a partir de uma sequência de caminhar.....	26
Figura 2.5: (a) Imagem com ruídos. (b) Imagem após filtragem Mediana.	27
Figura 2.6: Exemplo de imagens originais de um ciclo de caminhar: do início até o fim de um ciclo de (a) até (f).....	30
Figura 2.7: Sequência do caminhar de um indivíduo representado pela silhueta cheia. Início e fim de um ciclo de (a) até (f).	31
Figura 2.8: Sequência do caminhar de um indivíduo representado pelo contorno da silhueta. Início e fim de um ciclo de (a) até f).	31
Figura 2.9: Vetor de distância da silhueta ao longo de vários ciclos de caminhar de duas pessoas distintas. (a) Pessoa 1. (b) Pessoa 2.....	32
Figura 3.1: Representação dos espaços vetoriais da PCA e da LDA.	36
Figura 4.1: Dois exemplos de topologias para HMM's. O modelo (b) foi utilizado neste trabalho.	44
Figura 5.1: Imagens das bases de dados CASIA A (a) e B (b) e (c).	54
Figura 5.2: Imagem da base de dados GATECH.	54
Figura 5.3: Gráfico da variância acumulada em função do número de componentes principais para a base de dados CASIA A.	56
Figura 5.4: Gráfico da variância acumulada em função do número de componentes principais para a base de dados CASIA B.....	56
Figura 5.5: Gráfico da variância acumulada em função do número de componentes principais para a base de dados GATECH.	57

Figura 5.6: Gráfico da taxa de acerto pelo método LDA em função do número de componentes discriminatórias.	60
Figura 5.7: Gráfico da taxa de acerto pelo método PCA-LDA para dpca igual a 100, 150 e 200, base de dados CASIA A.	62
Figura 5.8: Gráfico da taxa de acerto pelo método PCA-LDA para dpca igual a 300, 400 e 300, base de dados CASIA A.	62
Figura 5.9: Gráfico da taxa de acerto pelo método PCA-LDA para dpca igual a 100, 150 e 200, base de dados CASIA B.	63
Figura 5.10: Gráfico da taxa de acerto pelo método PCA-LDA para dpca igual a 300, 400 e 500, base de dados CASIA B.	63
Figura 5.11: Gráfico da taxa de acerto pelo método PCA-LDA para dpca igual a 40 e 50, base de dados GATECH.	64
Figura 5.12: Gráfico da taxa de acerto pelo método PCA-LDA para dpca igual a 60 e 70, base de dados GATECH.	64
Figura 5.13: Gráfico da taxa de acerto pelo método POV, base de dados GATECH.	68

Lista de Tabelas

Tabela 5.1: Taxa de acerto (%) pela abordagem PCA para a base de dados CASIA A.....	58
Tabela 5.2: Taxa de acerto (%) pela abordagem PCA para a base de dados CASIA B.....	58
Tabela 5.3: Taxa de acerto (%) pela abordagem PCA para a base de dados GATECH.	58
Tabela 5.4: Limiar do coeficiente POV e redução do número de linhas dos vetores, para a base de dados CASIA A.	67
Tabela 5.5: Limiar do coeficiente POV e redução do número de linhas dos vetores, para a base de dados CASIA B.	67
Tabela 5.6: Limiar do coeficiente POV e redução do número de linhas dos vetores, para a base de dados GATECH.....	67

Capítulo 1

Introdução

1.1 Motivação

A sociedade moderna tem buscado sistemas computacionais cada vez mais sofisticados que possam oferecer sistemas seguros e confiáveis. Aeroportos, metrô, bancos, áreas de acesso público em geral são os principais interessados por métodos de identificação de indivíduos isentos de falhas ou que possibilitem o máximo de segurança por aqueles que nestes ambientes circulem (Nikolaos et al., 2005).

Os métodos tradicionais de segurança utilizados são ultrapassados e não suprem as necessidades dos sistemas de segurança. Por esta razão, o desenvolvimento e a utilização de métodos de autenticação através das características biométricas têm atraído a atenção das agências governamentais dentre outras instituições. Os principais métodos de reconhecimento computacional utilizam as características biométricas, que são características físicas e comportamentais de pessoas em geral (Jain, 2004). As mais comumente utilizadas são: face, impressão digital, íris e voz.

O reconhecimento biométrico refere-se ao reconhecimento automático de indivíduos utilizando uma ou mais características biométricas que são capturadas e processadas de forma que possam ser utilizadas como um meio de identificação de indivíduos. Esta identificação, dependendo da aplicação, vai permitir ou não o acesso da pessoa a uma determinada área ou região, ou somente fazer seu reconhecimento. Ou seja, existem sistemas de verificação e identificação.

No sistema de verificação, é validada a identificação da pessoa comparando a característica biométrica capturada com suas amostras armazenadas num sistema de banco de dados (problema 1 para 1). Neste caso, é necessário uma identificação convencional como um cartão magnético, o nome da pessoa ou uma senha e o sistema em seguida faz a comparação da característica biométrica daquele indivíduo com sua amostra na base de dados. A identificação por verificação normalmente é utilizada quando se quer evitar múltiplas pessoas

com a mesma identidade, de forma a permitir apenas um indivíduo. Já no sistema de identificação o sistema faz a captura da característica biométrica do indivíduo e a compara com as amostras de todos os indivíduos no banco de dados, estabelecendo uma identificação individual (problema 1 para N). Neste caso não é necessário se fazer uma identificação prévia (Jain et al., 2004)

Neste trabalho, será mostrada uma característica biométrica que tem ganhado destaque nas mais recentes pesquisas da área de visão computacional, a forma de caminhar (do inglês GAIT). A forma de caminhar é uma característica biométrica muito peculiar de cada indivíduo, pois cada pessoa tem a sua forma física, ou seja, possui uma silhueta distinta do outro e a dinâmica com que a pessoa caminha também varia de indivíduo para indivíduo. Um exemplo é quando se observa uma pessoa de costas caminhando na rua. Neste caso, sem se ver a sua face, ainda assim, consegue-se identificar quem é esta pessoa, observando-se apenas o seu caminhar. Estudos em psicofísica indicam que os humanos possuem a capacidade de reconhecer pessoas através do caminhar, mesmo tendo uma visão não tão nítida da imagem, indicando então a existência de informações relevantes na forma de caminhar (REF). Estudos atuais na área de medicina indicam que existem aproximadamente vinte e quatro componentes diferentes no caminhar humano e que o caminhar é único em cada indivíduo (Kale et al., 2004). Por isso, esta característica biométrica tem a vantagem de ser mais imune a falsificação.

Em relação às outras características biométricas mais comuns, a forma de caminhar possui como vantagem o fato de poder ser capturada à uma distância considerável, através de uma câmera e não necessitar da colaboração direta do indivíduo em questão (Nikolaos et al., 2005). Isto possibilita que o sistema não necessite de instrumentos digitais tão sofisticados e caros, tornando-o acessível e de baixo custo. A maioria das outras características biométricas como face, impressão digital, íris e voz, só podem ser capturadas através do contato físico direto ou a uma distância bem próxima ao sensor.

Entretanto, o reconhecimento pela forma de caminhar também possui algumas limitações. Sua identificação pode ser comprometida pela roupa que o indivíduo estiver vestindo, se esta for volumosa, por uma bolsa que estiver carregando em mãos ou nas costas, por calçados se forem muito diferentes do habitual, pela forma do cabelo, se este sofreu modificação, entre outros exemplos. Um ponto sensível é se a pessoa, por algum motivo, sofreu algum tipo de lesão corporal causado por acidente ou doença adquirida, e

consequentemente teve seu estilo de andar alterado. Nesta situação, o sistema de reconhecimento poderá não identificar este indivíduo de forma correta, sendo mais prudente neste caso mudar o sistema de identificação deste indivíduo (Salomão et al., 2007). Uma forma alternativa seria utilizar o caminhar como uma característica biométrica discriminante ou em conjunto com outra característica, fazendo-se a fusão destas. Neste sentido, sistemas biométricos multimodais se tornam muito atrativos e vêm sendo estudados em muitos trabalhos atuais, como em (Salomão et al., 2007). A idéia principal é fazer uma associação entre uma característica biométrica já muito comum e que possua alta taxa de acertos no reconhecimento, como voz e face, com mais uma característica, como o caminhar, que possa ser utilizada de forma discriminante. Por isso da importância de se estudar novas formas biométricas e aprimorar suas aplicações.

Um sistema de identificação de indivíduos abrange várias áreas, como reconhecimento de padrões, processamento de sinais e imagens, segmentação, visão computacional, inteligência artificial, entre outras. Estas áreas de concentração permitem, através das suas especificidades, processar as características biométricas capturadas por sensores diversos, entre elas a forma de caminhar, e utilizá-las para classificação de pessoas, e assim aplicar estes sistemas nas mais diversas áreas de segurança.

1.2 Trabalhos Relacionados

O estudo do caminhar iniciou na década de 70, mas somente a partir da década de 90, conseguiu-se obter resultados expressivos na área de visão computacional. As pesquisas se dividiram em dois grandes grupos, o primeiro grupo se refere aquelas em que o caminhar foi aproximado para modelos representativos da forma física ou da silhueta e o outro grupo pela análise da dinâmica corporal, sem um modelo específico.

Nos trabalhos de (Cunado et al., 1997), o caminhar foi modelado como um pêndulo articulado e foi utilizada a Transformada de Hough para extrair uma linha representativa da coxa do indivíduo em cada quadro de uma sequência de vídeo. A Análise de Fourier foi aplicada na inclinação extraída da coxa gerando espectro de magnitudes fase-peso de onde se formou as assinaturas para o reconhecimento. Já (Johnson et al., 2001) usaram parâmetros estáticos do corpo humano durante o caminhar sobre múltiplos ângulos de visão, a fim de se

obter uma representação tri-dimensional da pessoa. Neste mesmo trabalho, foi desenvolvida uma métrica que permitia prever quão bom um conjunto de vetores extraídos poderiam filtrar identidades numa grande população. Em (Yam et al., 2002) foi realizada uma comparação entre o andar e o correr de uma pessoa, calculando a inclinação da coxa e da perna e, a partir destas informações, aplicou-se também a Transformada de Fourier resultando em uma assinatura. Uma abordagem foi totalmente automática e outra não automática requeria especificação de um parâmetro que distinguisse andar de correr. O método mostrou-se tolerante a ruídos e a baixa resolução.

A efetividade dos resultados da abordagem por modelos, especialmente quando são utilizados a estrutura do corpo e parâmetros, são limitados por permitirem técnicas de visão imperfeitas (a localização do corpo através de espaços bi-dimensional e tri-dimensional tem sido um desafio e um problema ainda não resolvido). Assim, o custo computacional se torna relativamente alto.

Considerando-se a abordagem por movimento, pode-se fazer o uso de poses estáticas do corpo e reconhecê-los considerando variações temporais das observações a estas poses, como em (Kale et al., 2002) e (Wang, et al., 2002). Ou ainda utilizar métodos de espaço-tempo, em geral caracterizada por distribuições de espaço-tempo geradas pelo caminhar em sua continuidade, como em (Little e Boyd, 1998), (Hayfron-Acquah et al., 2002) e (Phillips et al., 2002).

Alguns trabalhos, como o de (Kale et al., 2004), utilizam a silhueta humana como vetor de características de duas formas, extraíndo apenas o contorno da silhueta e medindo a largura desta ou extraíndo a silhueta cheia, mas aumentando consideravelmente a dimensão do vetor. A extração do contorno da silhueta é uma forma de se reduzir a dimensionalidade dos dados, sem perder a forma e moldura do corpo, mas tentando manter, em certa medida, a dinâmica do caminhar, já que as silhuetas são obtidas a partir de sequências de imagens (Wang e Tan, 2003). Desta forma, o vetor de característica é formado por sequências de silhuetas que variam ao longo dos quadros do vídeo capturado.

Em (Salomão et al., 2006a) e (Salomão et al., 2006b), empregou-se a técnica ESC (Energia das Silhuetas do Caminhar) para a formação dos vetores de características. Tal técnica foi adaptada da abordagem conhecida como Imagem da Energia do

Movimento/Imagem do Histórico do Movimento (*motion-energy image/motion-history image - MEI/MHI*) (Bobick e Davis, 2001) (Han e Bhanu, 2006), onde se projeta as informações de uma sequência de caminhar em uma única imagem. Desta forma, utilizando a silhueta cheia e armazenando seus níveis de energia em nível de cinza, buscou-se preservar as informações temporais da forma de caminhar.

Entretanto, a projeção de todos os quadros de um vídeo de movimento em uma única imagem, mesmo codificando-se parte da informação nos níveis de cinza da imagem criada (técnica ESC), pode não ser suficiente para se avaliar, de maneira adequada, a dinâmica do problema. A fim de se capturar a dinâmica da caminhada, acredita-se que abordagens estocásticas sejam mais adequadas que abordagens estáticas com extração estatísticas de parâmetros. Neste sentido, em (Kale et al., 2004) e (Liu e Sarkar, 2006) foi abordado o uso de modelagem dinâmica, por meio de Modelos Ocultos de Markov (*Hidden Markov Models - HMM*), na identificação da forma de caminhar, mostrando bons resultados. O HMM pode se relacionar à característica biométrica do caminhar, associando-se a transição de estados do modelo aos estados presentes num ciclo de caminhar. Estes estados são as poses típicas de uma pessoa quando esta está caminhando, e que ocorrem ciclicamente. Outro ponto importante está na dinâmica presente em cada indivíduo quando anda, acreditando-se que esta é distinta em cada pessoa.

Um grande número de pesquisas com as mais diversas abordagens têm mostrado que é possível reconhecer pessoas pela forma de caminhar. A maior parte das abordagens utiliza capturas frontais e paralelas dos indivíduos de uma câmera estática, com um fundo plano e sem oclusão. Comparando-a com as outras características biométricas mais largamente utilizadas como face, voz e impressão digital, o reconhecimento pela forma de caminhar ainda está na fase de “infância” (Wang et al., 2003).

1.3 Objetivos da Pesquisa

Diante das poucas pesquisas existentes (de ser um tema recente) na área de reconhecimento de indivíduos pela forma de caminhar, este tema se torna muito interessante e com muito a ser explorado. A maior parte dos sistemas de reconhecimento emprega as características biométricas mais tradicionais, entretanto estas possuem algumas limitações,

como a necessidade da colaboração do indivíduo ou mesmo do contato direto com o objeto de captura. Por isso é de suma importância estudar e desenvolver alternativas que possam suprir estas necessidades.

Desta forma, este trabalho tem como principal objetivo estudar como é composto o caminhar de indivíduos e avaliar um modelo dinâmico que tenha a capacidade de absorver as informações da dinâmica de cada indivíduo ao longo do tempo, a fim de desenvolver um sistema automático de identificação de indivíduos. A partir deste estudo, é possível analisar quão eficiente e confiável o sistema de reconhecimento pela forma de caminhar pode ser e se o mesmo pode aprimorar sistemas de reconhecimento multimodais.

Serão mostradas algumas técnicas tradicionais de redução de dimensionalidade e classificação na área de reconhecimento de voz e face, com a finalidade de verificar se estes se aplicam satisfatoriamente ao caminhar.

1.4 Organização do Texto

O trabalho foi dividido em capítulos. O capítulo presente explica o que é biometria e como pode ser feita a sua utilização na área de segurança. Em seguida foi feita uma descrição das principais pesquisas na área de reconhecimento pela forma de caminhar, explicitando as abordagens mais utilizadas, como as por modelos pré-definidos e as por dinâmica corporal. Ao final foram explicitados os objetivos do trabalho.

No Capítulo 2 é explicado em que consiste o caminhar de um indivíduo, justificando, assim, a escolha da abordagem testada. É descrito, também, um método de extração de características robusto, de onde é estimando o fundo das sequências de vídeo para realizar a segmentação das silhuetas e desta forma gerar os vetores de dados que sejam de qualidade aceitável. O Capítulo 3 descreve os métodos de redução dimensional das características extraídas, a fim de diminuir o custo computacional na etapa de classificação sem perder a representatividade dos dados. No Capítulo 4 é apresentado o classificador testado, suas principais características, entre elas a de possuir uma alta capacidade de representar e modelar as mudanças contidas nas características estatísticas das observações, e a justificativa de sua aplicação.

O Capítulo 5 descreve detalhadamente os procedimentos adotados nas duas etapas de testes do sistemas de reconhecimento desenvolvido. A primeira etapa consistiu em estudar e avaliar sistemas de reconhecimento de indivíduos utilizando uma base de dados de domínio público, com ambiente controlado, e com grande número de indivíduos (100 no total), seja o indivíduo caminhado normalmente seja ele caminhando usando roupas distintas e carregando uma bolsa. A escolha de uma base de dados com ambiente controlado teve a finalidade de, primeiramente, observar se o classificador responderia bem as características extraídas, levando-se em conta a dinâmica do caminhar, sem haver uma grande preocupação com a parte de segmentação das imagens. A segunda etapa consistiu na utilização de outra base de dados também de domínio público, com um número reduzido de pessoas (16 no total), mas com um ambiente não tão bem controlado. Este fato exigiu maior rigor e preocupação na segmentação das silhuetas, devido à iluminação do ambiente em questão, que não era tão boa, e às cores das vestes dos indivíduos que dificultaram a segmentação. Estes dois testes são importantes, pois indicaram o quão robusto é o sistema de reconhecimento desenvolvido. Observe que, na prática, os sistemas de segurança podem ter variações nos ambientes em que se aplicam e as pessoas a serem identificadas podem estar com roupas que não colaborem para a extração de características. Estes aspectos não podem ser motivos para que ocorram falhas no sistema.

No Capítulo 6 é feita uma conclusão do trabalho, mostrando a sua contribuição para a comunidade de reconhecimento de indivíduos e sugestões para trabalhos futuros.

Capítulo 2

Extração do Vetor de Características

2.1 Introdução

O reconhecimento humano através da forma de caminhar pode ser realizado de acordo com o diagrama ilustrado na Figura 2.1. De forma geral, a primeira etapa abrange a extração de características no qual é utilizado um sensor (podendo ser uma câmera analógica ou digital, ou até mesmo uma câmera infravermelha) para captar as imagens do andar humano. Capturadas as imagens, é feita uma segmentação de imagens e extração de silhuetas para formação do vetor de características. Em seguida é aplicado um redutor dimensional com o propósito de aproveitar apenas aquelas informações relevantes que compõem a assinatura da forma de caminhar e diminuir o esforço computacional. Na última etapa é aplicado o classificador a fim de que o sistema identifique qual é o indivíduo para um dado conjunto de vetores de características obtidos na etapa 2.

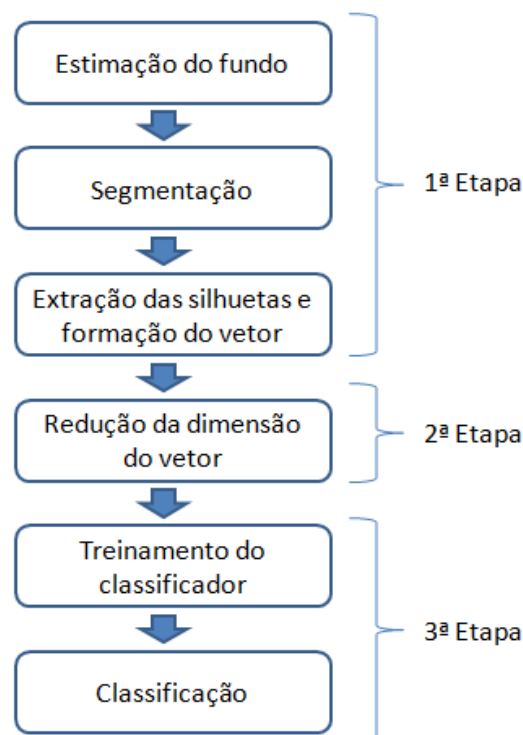


Figura 2.1: Diagrama de um sistema de identificação de indivíduos.

A extração de características é a primeira etapa para a formação do vetor de características no sistema de reconhecimento de indivíduos proposto. A qualidade da extração da silhueta é de grande importância, pois se esta etapa é feita com qualidade, garante-se que as informações mais relevantes serão armazenadas e aproveitadas. Caso contrário, pode-se comprometer os resultados finais refletindo numa redução da taxa de reconhecimento. Desta forma, a segmentação das imagens das sequências de vídeo deve ter um tratamento específico para que se eliminem possíveis ruídos, e assim as silhuetas possam representar fielmente cada um dos indivíduos da base de dados.

2.2 A Forma de Caminhar

A forma de caminhar pode ser vista como uma atividade cíclica, periódica, composta por dois passos, na qual em cada ciclo são mantidas a média e a variância das silhuetas extraídas. Como cada indivíduo tem uma forma de caminhar única, ou seja, distinta entre indivíduos, e assumindo-se algum método de caracterização de silhuetas, pode-se considerar que as variáveis em questão são aleatórias, tendo-se então um processo estocástico. As informações contidas na forma de caminhar podem ser pré-processadas através de técnicas de processamento de imagens e reconhecimento de padrões.

A análise do caminhar deve utilizar características razoavelmente robustas para que operem de forma discriminatória entre as pessoas. Por isso, é utilizado como característica a silhueta de indivíduos, que intuitivamente parece ser uma boa representação da forma e do movimento do corpo humano (Kale et al., 2004). Pode-se considerar duas abordagens para a utilização da silhueta de indivíduos como característica na forma de caminhar, a abordagem por modelos e a abordagem holística (Nikolaos et al., 2005) (Cheng et al., 2007). As Figuras 2.2 e 2.3 ilustram alguns exemplos destas duas abordagens:

- a) Abordagem por modelos: Esta abordagem utiliza modelos pré-definidos que se aproximem de uma silhueta humana, utilizando formas geométricas semelhantes aos membros do corpo humano. Costuma-se empregar conjuntos de elipses, circunferências, retângulos ou a combinação destas, medidas de ângulos destes membros e características estáticas como a altura, distância entre a cabeça e a pélvis, distância entre os pés, entre outras. Como neste caso é feita a identificação de pontos

específicos do corpo, é necessário que se capture as imagens da sequência de vídeo com alta qualidade para que sejam válidos.

- b) Abordagem holística: A abordagem holística opera diretamente na sequência do caminhar do indivíduo, sem assumir modelos específicos. Utiliza-se a silhueta em si como característica e tem como vantagem a possibilidade de se analisar a dinâmica de cada indivíduo numa sequência de imagens. Os principais tipos de medidas utilizadas a fim de acompanhar a dinâmica associada ao contorno da silhueta são medidas de largura da silhueta, suas projeções horizontais e verticais e representações angulares. As informações de cores e texturas não são informações importantes para garantir a captura da dinâmica, apenas a extração da silhueta de forma binária (preto e branco) se faz necessária.

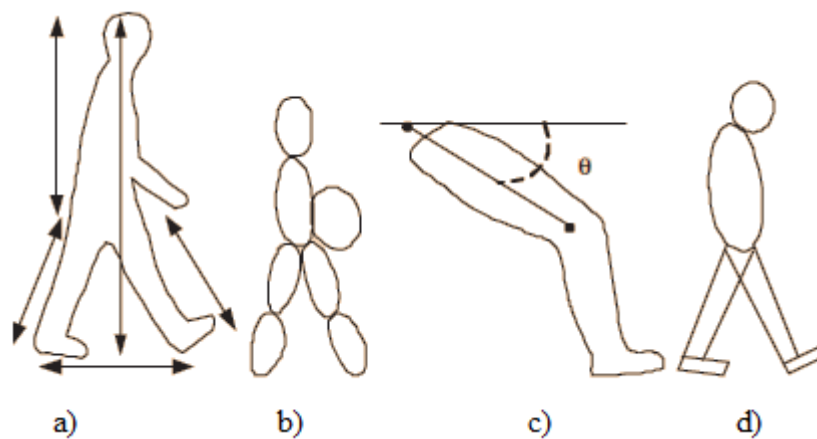


Figura 2.2: Características extraídas de silhuetas binárias a partir da abordagem por modelos.

(a) Medida de distâncias, (b) elipses, (c) angulação da perna, (d) combinação de formas geométricas (Nikolaos et al., 2005).

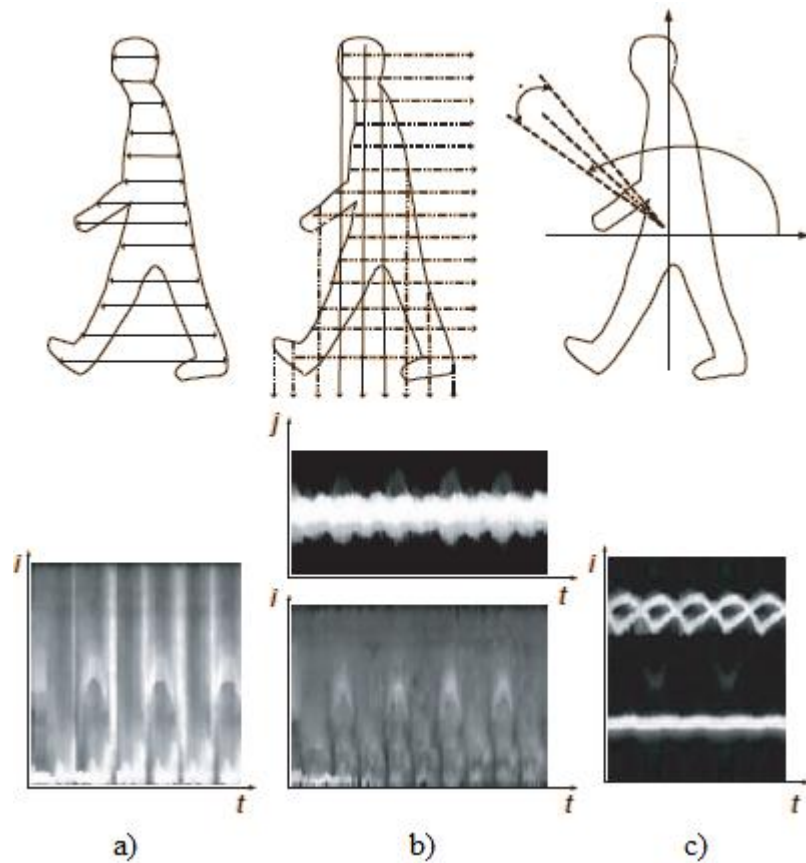


Figura 2.3: Características extraídas de silhuetas binárias a partir da abordagem holística, em relação ao tempo: (a) largura da silhueta, (b) projeção vertical e horizontal, (c) representação angular (Nikolaos et al., 2005).

2.3 Extração dos Dados da Sequência de Vídeo

A extração da silhueta é feita, primeiramente, a partir da captura do vídeo do caminhar, pela extração do fundo da cena, fundo convenientemente estimado. A extração do fundo é obtida efetuando-se a subtração, para cada quadro de vídeo, pelo fundo estimado, obtendo-se apenas a silhueta do indivíduo. A técnica utilizada para a obtenção do fundo, conhecida como LMedS (*Least Median Square*), realiza um procedimento de estimação robusta para a sua construção, utilizando as mesmas sequências de imagens que serão empregadas como dados (Wang e Tan, 2003) (Stewart, 1999).

Seja uma sequência de N imagens representada por I . O fundo b_{xy} pode ser obtido por

$$b_{xy} = \min_p \text{med}_t (I_{xy}^t - p)^2 \quad (2.1)$$

onde p é a intensidade do brilho do pixel na posição (x,y) , med é a mediana calculada e t representa o índice do quadro do vídeo (Salomão et al., 2005). Um número de aproximadamente 50 a 60 quadros é suficiente para gerar o fundo. A Figura 2.4 ilustra um exemplo de um quadro de uma sequência de imagens e o seu fundo estimado.



Figura 2.4: (a) Quadro de uma sequência de caminhar. (b) Fundo estimado a partir de uma sequência de caminhar.

Depois de obtido o fundo é necessário fazer a subtração pixel a pixel de cada sequência de imagem com o seu fundo correspondente para a obtenção da silhueta. Esta subtração pode ser realizada diretamente, através da própria operação matemática, quando se têm imagens capturadas em locais de iluminação controlada, por exemplo, locais internos e fechados, ou pode empregar uma função de compensação de brilho de acordo com a equação a seguir (Wang e Tan, 2003):

$$f(a,b) = 1 - \frac{2 \times \sqrt{(a+1)(b+1)}}{(a+1)(b+1)} \times \frac{2 \times \sqrt{(256-a)(256-b)}}{(256-a)(256-b)} \quad (2.2)$$

para $0 \leq f(a,b) < 1$ e $0 \leq a(x,y), b(x,y) \leq 255$,

onde $a(x,y)$ e $b(x,y)$ são a intensidade do brilho da imagem e do fundo na posição (x,y) , respectivamente, e $f(a,b)$ é a função de extração. Quando a é igual a b , o valor da função de extração $f(a,b)$ é igual ao valor mínimo, zero. Mas, conforme a diferença entre a e b aumenta, a função de extração $f(a,b)$ também aumenta (Kuno et al., 1996). Essa

compensação de brilho visa diminuir os efeitos de iluminação nas imagens (por exemplo, imagens capturadas em ambientes externos ou sob forte influência da iluminação local, o que pode causar algum tipo de distorção ou irregularidade nas silhuetas extraídas).

Durante o procedimento anterior, observou-se uma contaminação por ruídos ao estilo sal e pimenta, ou seja, ruídos impulsivos aleatórios de alta amplitude que geram “chuviscos” na imagem. Portanto, para amenizar possíveis ruídos nas imagens obtidas após a subtração do fundo, aplicou-se um filtro Mediana, dando como resultado imagens mais suavizadas, porém sem perder a nitidez da imagem. O filtro Mediana é uma alternativa quando se tem o objetivo de reduzir ruídos sem borrar a imagem, principalmente se esta possui bordas a serem consideradas. O nível de cinza de cada pixel é substituído pela mediana dos níveis de cinza na vizinhança daquele pixel (Gonzalez e Woods, 2002).

A Mediana m de um conjunto de valores é tal que metade dos valores no conjunto são menores do que m e metade dos valores são maiores que m . Para calcular a filtragem por Mediana em uma vizinhança de pixel de uma imagem, primeiramente, seleciona-se os valores do pixel e de seus vizinhos, determina-se a mediana e atribui-se este valor ao valor do pixel. Por exemplo, em uma vizinhança 3×3 , a mediana é o 5º maior valor, em uma vizinhança 5×5 é o 13º maior valor, e assim por diante. Quando vários valores em uma vizinhança forem iguais, todos esses valores iguais devem ser agrupados. Assim, a função da filtragem Mediana é forçar pontos com intensidades distintas, assemelhem-se aos seus vizinhos, eliminando-se “chuviscos” de intensidades que apareçam isolados na área da máscara do filtro. A máscara utilizada no algoritmo descrito deste trabalho foi de 3×3 . A Figura 2.5 ilustra o efeito resultante da Filtragem Mediana sobre uma imagem que continha ruídos.



Figura 2.5: (a) Imagem com ruídos. (b) Imagem após filtragem Mediana.

Ao extrair a silhueta humana, pode-se escolher duas opções, utilizar a silhueta cheia ou apenas o seu contorno. A escolha por uma das duas opções está relacionada à qualidade das silhuetas binárias. Se as silhuetas possuem boa qualidade, o contorno possui todas as informações necessárias para sua representatividade, além de gerar uma dimensão bem reduzida em relação à silhueta cheia. Mas se a silhueta não for extraída com boa qualidade o contorno da silhueta provavelmente pode não ser fiel à imagem original. Neste caso, as imagens de silhuetas extraídas possuem boa qualidade, considerando-se que a maior parte das imagens de silhueta não obtiveram problemas de representatividade.

Como se quer apenas o contorno da silhueta e não ela cheia, empregou-se também uma filtragem de Canny para detecção de bordas (transição do nível de cinza escuro para o claro). A filtragem de Canny é basicamente uma filtragem para detectar bordas olhando para máximos locais do gradiente da imagem em questão. O gradiente é calculado usando a derivada de um filtro Gaussiano. De uma maneira geral, é requerido de todo algoritmo de detecção de bordas uma baixa taxa de erro na sua identificação. Igualmente, espera-se que os pixels das bordas sejam localizados com precisão, minimizando a distância entre os pixels encontrados pelo detector e a borda verdadeira. Um terceiro critério na detecção deveria evitar a possibilidade da geração de múltiplas bordas para expressar uma borda simples. E é baseado nestes aspectos que a filtragem tipo Canny foi desenvolvida, pois ela utiliza um detector de bordas ótimo para imagens com borda do tipo degrau, afina as bordas pelo processo chamado de supressão não máxima e elimina a fragmentação das bordas causadas por ruídos pelo processo chamado de histerese (Canny, 1986).

Um limiar de binarização é também aplicado para retirar algumas minúcias que não puderam ser filtradas. Mesmo após estes processos de filtrações, as silhuetas podem ainda apresentar espaços incompletos (buracos), sendo necessário o uso de operadores morfológicos. A morfologia é uma ferramenta para a extração de componentes de imagens que sejam úteis na representação e descrição da forma de uma região, como fronteiras, esqueletos e o fecho convexo. A linguagem da morfologia matemática é a teoria de conjuntos. Foram utilizados dois operadores morfológicos, a dilatação e a erosão. Mas antes de discutir o que são estes dois operadores, têm-se algumas definições básicas de translação, reflexão, complemento e diferença.

Assim, sejam A e B conjuntos de Z^2 , com componentes $a = (a_1, a_2)$ e $b = (b_1, b_2)$, respectivamente. A translação de A por $x = (x_1, x_2)$, denotada por $(A)_x$, é definida como (Gonzalez e Woods, 2002)

$$(A)_x = \{c | c = a + x, a \in A\}. \quad (2.3)$$

A reflexão de B, denotada por \hat{B} , é definida como

$$\hat{B} = \{x | x = -b, b \in B\}. \quad (2.4)$$

O complemento do conjunto A é definido como

$$A^C = \{x | x \notin A\}. \quad (2.5)$$

A diferença entre dois conjuntos A e B, denotada por A-B, é definida como

$$A - B = \{x | x \in A, x \notin B\}. \quad (2.6)$$

Portanto, a dilatação de A por B é a reflexão de B em torno de sua origem, seguido da translação desta reflexão por x, ou seja,

$$A \oplus B = \left\{ x \mid (\hat{B})_x \cap A \neq \emptyset \right\}. \quad (2.7)$$

Outra forma de interpretação do processo de dilatação de A por B é o conjunto de todos os deslocamentos x tais que \hat{B} e A sobreponham-se em pelo menos um elemento não nulo.

A erosão de A por B é o conjunto de todos os pontos x tais que B, quando transladado por x, fique contido em A, ou seja,

$$A \ominus B = \left\{ x \mid (B)_x \subseteq A \right\}. \quad (2.8)$$

A dilatação tem a função de completar as silhuetas nos pontos em que ficaram incompletas, e a erosão de afinar as linhas que foram levemente alargadas (Salomão et al., 2005).

Nas Figuras 2.6, 2.7 e 2.8, pode-se observar um exemplo de um ciclo de caminhar de uma pessoa, mostrando a imagem original, as suas silhuetas cheias e o contorno de suas silhuetas, respectivamente.

Para a formação do vetor de características de cada um dos indivíduos a serem classificados, foi utilizada a abordagem holística, visto que esta pode, ao mesmo tempo, associar as silhuetas de forma estática, mas observando suas variações ao longo de um ciclo de caminhar. Desta forma, as informações da dinâmica do caminhar não são perdidas. Portanto, é calculada a largura da silhueta, ou seja, a distância entre a borda mais a esquerda e a borda mais a direita da silhueta, linha por linha ao longo de uma sequência de imagens. Estas medidas de distâncias colocadas sobre a forma de um vetor podem ser vistas na Figura 2.9.

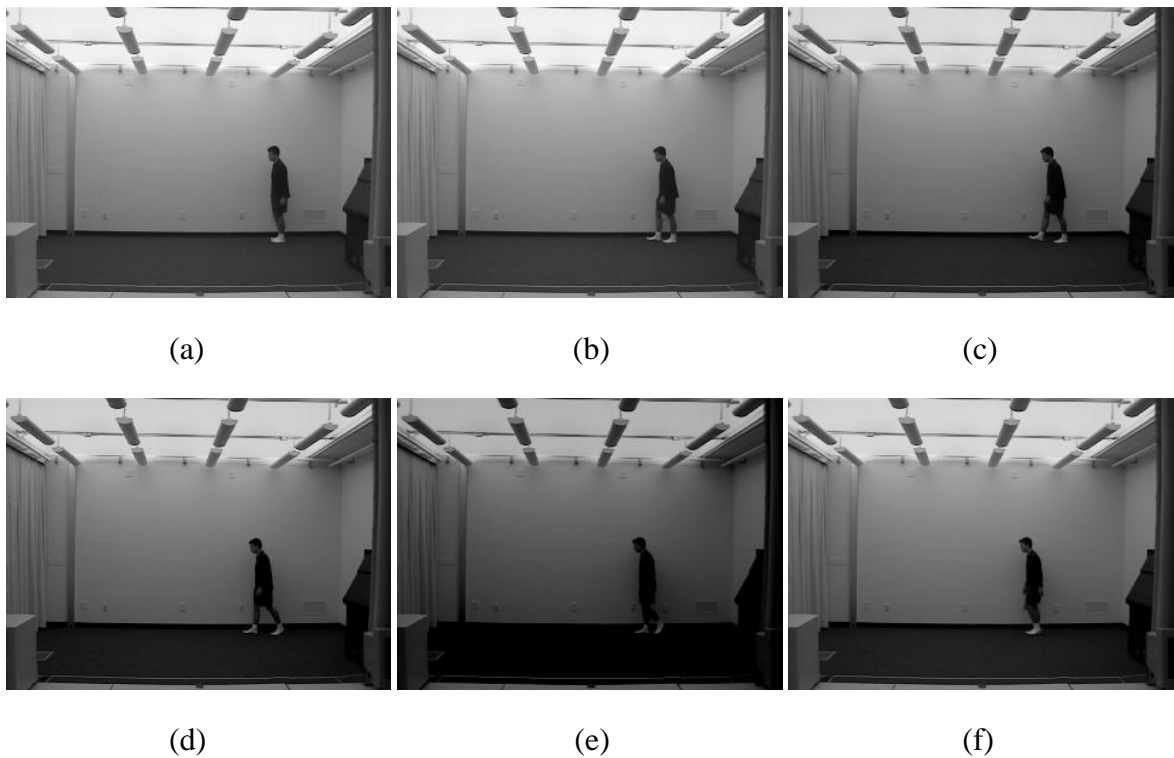


Figura 2.6: Exemplo de imagens originais de um ciclo de caminhar: do início até o fim de um ciclo de (a) até (f).

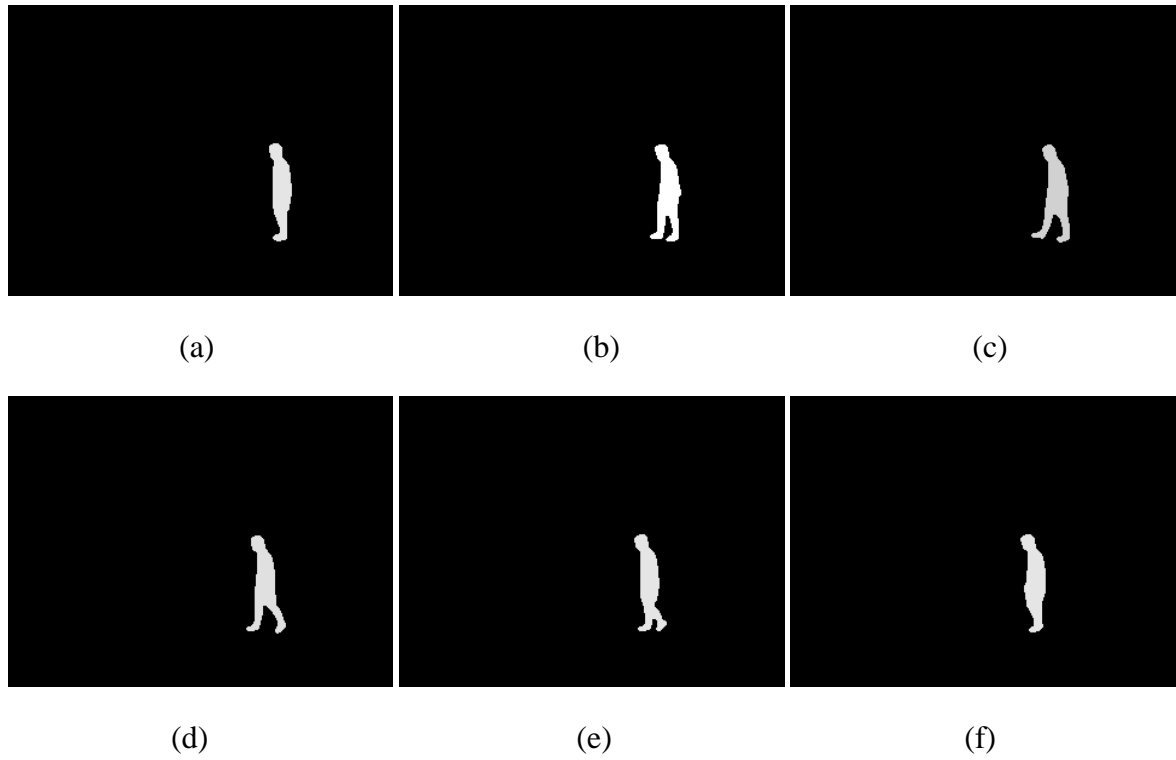


Figura 2.7: Sequência do caminhar de um indivíduo representado pela silhueta cheia. Início e fim de um ciclo de (a) até (f).

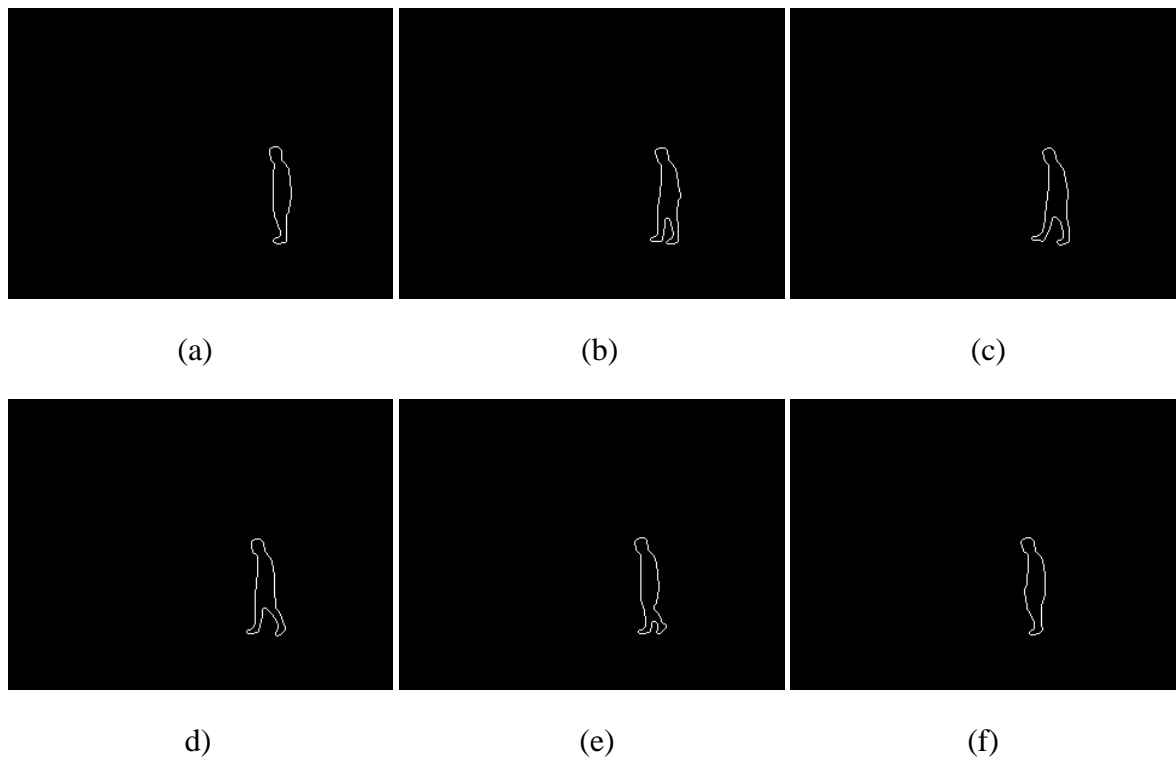


Figura 2.8: Sequência do caminhar de um indivíduo representado pelo contorno da silhueta. Início e fim de um ciclo de (a) até f).

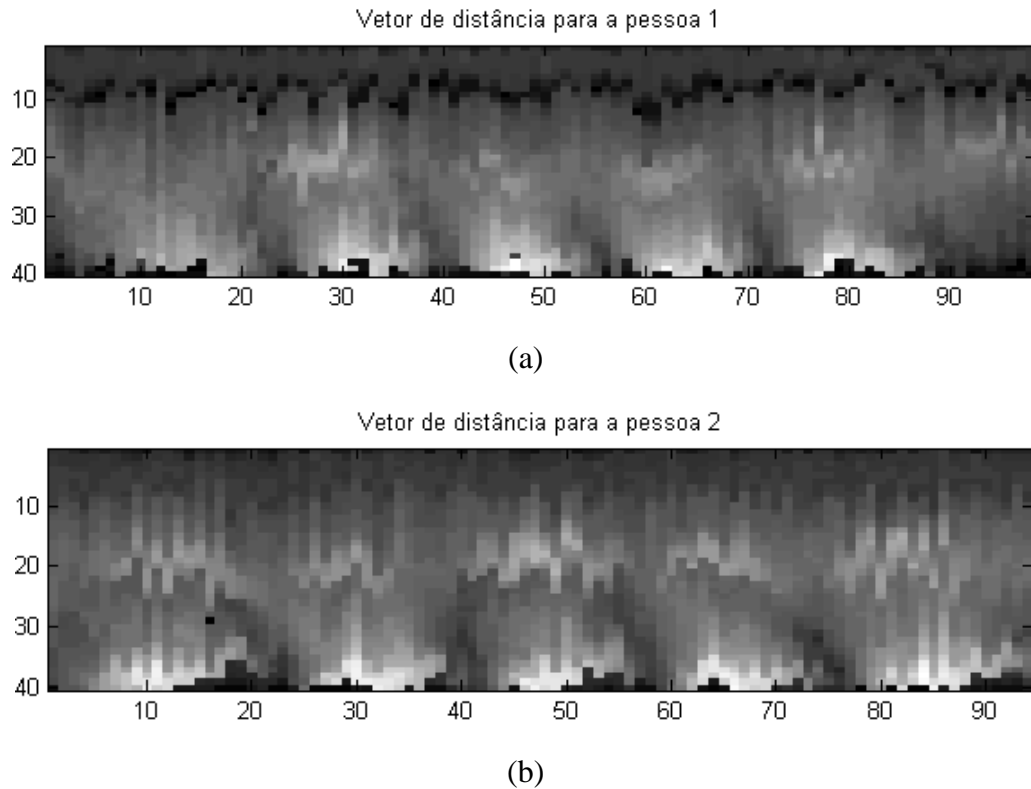


Figura 2.9: Vetor de distância da silhueta ao longo de vários ciclos de caminhar de duas pessoas distintas. (a) Pessoa 1. (b) Pessoa 2.

A partir das Figuras 2.9 (a) e (b), é possível observar o comportamento cíclico do caminhar de uma pessoa, e este comportamento é distinto para pessoas distintas. Considerando que cada silhueta seja representada por 40 medidas de distância, e que estas medidas são calculadas a cada sequência de vídeo, vemos que o gráfico possui no eixo vertical 40 pontos e no eixo horizontal o número de sequências de vídeo específico para cada indivíduo (variável).

Os pontos de cores mais claras mostram as partes do corpo humano que possuem uma maior distância entre as extremidades e os pontos mais escuros representam as partes mais próximas. Desta forma, vê-se que algumas partes do corpo como a cabeça, braços, quadril e pés, variam entre as cores branco e cinza escuro com o aumento da sequência de vídeo, pois são partes do corpo de maior dinâmica.

Capítulo 3

Técnicas de Redução de Dimensão

3.1 Introdução

Um fator determinante em um sistema de reconhecimento de padrões é a dimensão dos dados. E este é um fator que pode comprometer a tarefa de reconhecimento de padrões. Por exemplo, os humanos têm uma capacidade excepcional de identificar padrões em espaços de 1, 2 ou 3 dimensões. Contudo, em espaços de dimensão maior eles não conseguem identificar nem padrões simples. A necessidade de visualização dos dados (pelos humanos) de alta dimensão leva à utilização de algumas técnicas de redução de dimensão. Em muitas aplicações, numerosas características são obtidas a fim de assegurar uma classificação acurada das classes desconhecidas. Algumas vezes, o número de características deve ser reduzido antes de usar um classificador, pois o método de classificação utilizado pode se tornar computacionalmente e/ou analiticamente intratável (Souza, 1999).

O problema de redução de dimensionalidade acentua-se em espaços de alta dimensão e deve-se utilizar ferramentas que possam modelar os dados. Busca-se um espaço de dimensão menor, mas que ainda represente a informação contida nestes dados. É razoável pensar assim, pois a quantidade de dados, muitas vezes, é confundida com a qualidade das relações existentes entre os mesmos.

Na literatura existente (Fukunaga, 1990), a redução de dimensionalidade está intimamente ligada à capacidade dos algoritmos em utilizar o espaço das características para promover uma boa classificação. Portanto, é útil ou mesmo necessário primeiro fazer um pré-processamento no sentido de reduzir a dimensão das características (dimensão dos dados) para um tamanho manuseável, ajustando as informações originais o quanto for possível. É possível pensar que um fenômeno representado por um espaço de padrões de alta dimensão, e assim complexo, pode realmente ser governado por umas poucas variáveis (espaço de características).

Nas próximas seções serão mostradas algumas destas técnicas utilizadas no presente trabalho de reconhecimento de indivíduos pela dinâmica do caminhar.

3.2 Análise de Componentes Principais

Muitos dos algoritmos clássicos de processamento de sinal recorrem a métodos de descorrelação de dados. As vantagens de tais técnicas baseiam-se num princípio relativamente simples: ao descorrelacionar os dados, elimina-se parte da informação redundante em cada dimensão. Certos sinais, como é o caso de sinais de fala e imagens, que apresentam uma substancial correlação entre amostras ou “pixels” consecutivos, são os candidatos ideais para os algoritmos de descorrelação. Entre os benefícios resultantes é possível citar:

- os dados podem ser descritos de uma forma mais concisa;
- certas características escondidas dos dados podem ficar mais explícitas depois de transformadas.

Neste sentido, um dos métodos mais utilizados para descorrelação de dados é conhecido como Análise de Componentes Principais (*Principal Components Analysis* - PCA).

A Análise de Componentes Principais é um dos métodos estatísticos mais utilizados quando se pretende analisar dados multivariados. Segundo (Kitani e Thomaz, 2007), na atualidade, a técnica PCA, além de uma das mais antigas, é também uma das mais bem sucedidas técnicas de estatística multivariada. Sua aplicação é muito comum na área de reconhecimento facial, e por isso esta técnica se mostra atrativa para tratar as demais características biométricas, entre elas a que está sendo utilizada neste trabalho, o caminhar.

A técnica PCA permite transformar um conjunto de variáveis correlacionadas em um novo conjunto de variáveis não correlacionadas, tendo maior sucesso desde que a distribuição dos dados seja gaussiana. O objetivo desta técnica é verificar se existe um número reduzido de componentes principais que seja responsável por explicar uma proporção elevada da variação total associada ao conjunto de dados original (Zhou e Bhanu, 2006). Desta forma, é possível reduzir a dimensão do vetor de características sem ter que trabalhar em um espaço de características de dimensão elevada (Zao et al., 1998).

Seja um conjunto Ω , formado por M vetores (coluna) obtidos no algoritmo de extração de silhuetas, descrito por

$$\Omega = \{\Gamma_i, \dots, \Gamma_M\}, \quad (3.1)$$

onde $\Gamma_i \in \mathbb{R}^b$, $i=1, \dots, M$. O vetor médio \mathbf{X} deste conjunto de dados pode ser calculado por:

$$\mathbf{X} = \frac{1}{M} \sum_{n=1}^M \Gamma_n, \quad (3.2)$$

obtendo-se o vetor \mathbf{V}_i de média zero pode ser obtido segundo

$$\mathbf{V}_i = \Gamma_i - \mathbf{X}, \quad i=1, \dots, M. \quad (3.3)$$

Assim, para a redução da dimensionalidade, deseja-se obter uma matriz de transformação $\mathbf{W} \in \mathbb{R}^{b \times a}$, $a < b$, tal que a projeção de uma silhueta Γ_i seja dada por

$$\mathbf{O}_i = \mathbf{W}^T \mathbf{V}_i, \quad i=1, \dots, M, \quad (3.4)$$

onde $\mathbf{O}_i \in \mathbb{R}^a$ é o vetor projetado com dimensão menor que o vetor original. Para isto, define-se uma matriz de espalhamento (Belhumeur et. al., 1997).

$$\mathbf{S} = \sum_{k=1}^M \mathbf{V}_k \mathbf{V}_k^T. \quad (3.5)$$

Os vetores transformados, após a aplicação de (3.5), possuem uma nova matriz de espalhamento dada por $\mathbf{W}^T \mathbf{S} \mathbf{W}$. Na análise de PCA, encontrar o vetor ótimo equivale a encontrar uma matriz de projeção ótima \mathbf{W}^* , obtida pela maximização do determinante da matriz de espalhamento projetada, ou seja:

$$\mathbf{W}^* = \arg \max_{\mathbf{W}} |\mathbf{W}^T \mathbf{S} \mathbf{W}|. \quad (3.6)$$

A solução de (3.6) gera como resultado uma matriz \mathbf{W}^* composta por ' a ' vetores coluna, que são os autovetores de \mathbf{S} , associados aos ' a ' maiores autovalores. Os dados são então projetados neste novo subespaço, reduzindo sua dimensão, estando adequados para serem utilizados como vetores de características na etapa de classificação dos indivíduos.

3.3 Análise de Discriminantes Lineares

A técnica de redução dimensional apresentada anteriormente através da PCA, tem como resultado uma redução efetiva ou mesmo uma representação mais econômica dos dados que têm alta dimensionalidade (Martínez e Kak, 2001). Entretanto, sabe-se que não necessariamente a direção apontada pelos autovetores do PCA indica a melhor direção para fins de classificação. A Figura 3.1 ilustra como as informações referentes a duas classes podem estar no espaço da PCA. As distribuições de duas classes podem estar superpostas no espaço PCA, não sendo possível fazer distinção entre pontos das duas classes. Mas, em outro espaço estas informações podem ser melhor representadas, sem que haja superposição entre as classes.

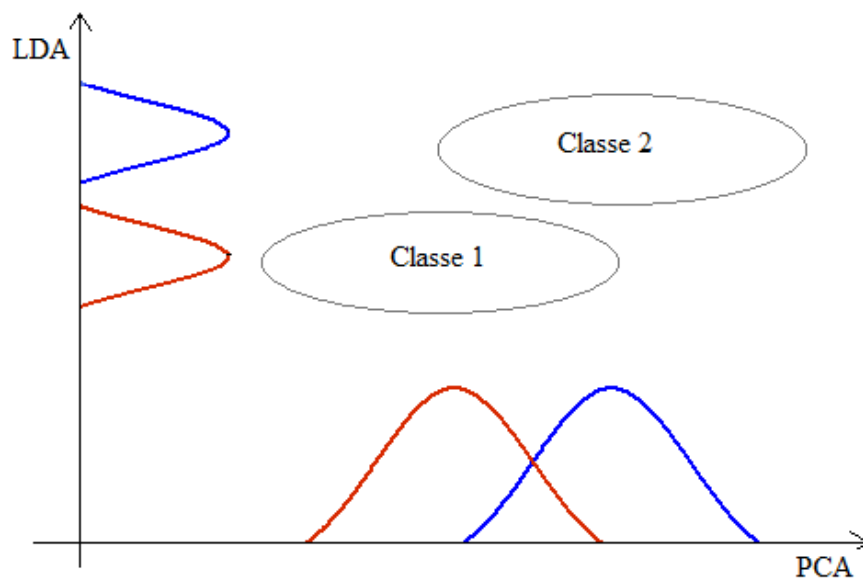


Figura 3.1: Representação dos espaços vetoriais da PCA e da LDA.

Desta forma, uma técnica complementar ou mesmo alternativa é a Análise de Discriminantes Lineares (*Linear Discriminant Analysis* – LDA). Esta técnica é baseada no princípio de se trabalhar com as projeções das amostras em uma base n -dimensional e determinar um subespaço vetorial que melhor discrimine as amostras entre classes, utilizando um critério de maximização da separação entre classes, que foi proposto por Ronald A. Fisher (Belhumeur et al., 1997). O subespaço de projeção dos dados é então escolhido de modo a que haja um aumento na dispersão entre as classes e uma diminuição na dispersão dentro das classes. Todas as vezes que um subespaço linear maximiza o quociente dispersão entre classes

e minimiza a dispersão dentro da classe, está-se diante de um discriminador linear de Fisher. Assim, a aplicação desta operação permite uma maior discriminação entre classes durante a etapa de classificação (Liu e Zheng, 2007). A Figura 3.1 faz esta comparação entre os espaços dimensionais PCA e LDA, para dois conjuntos de classes distintas.

Observe que uma das dificuldades do método é converter uma razão entre matrizes para um número, ou índice, que possa expressar o quão bom é o subespaço final obtido. Ou seja, é necessário definir uma forma que possa medir como, em um determinado subespaço, o espalhamento se verifica e, assim, decidir por ele ou por outro. De outra forma, este número precisa ser grande quando o espalhamento entre classes é grande e o espalhamento dentro das classes é pequeno. Diversos operadores são propostos na literatura (Fukunaga, 1990) como o traço de cada matriz, o determinante ou o logaritmo natural do determinante. Neste trabalho, optou-se pelo uso do determinante como medida de índice, como será visto a seguir.

Seja a matriz de variância entre-classes definida como

$$S_B = \sum_{i=1}^c N_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T \quad (3.7)$$

e a matriz de variância intra-classes definida como

$$S_W = \sum_{i=1}^c \sum_{\Gamma_k \in X_i} (\Gamma_k - \boldsymbol{\mu}_i)(\Gamma_k - \boldsymbol{\mu}_i)^T, \quad (3.8)$$

onde $\boldsymbol{\mu}_i \in \mathbb{R}^a$ é o vetor imagem média da classe Γ_i (silhueta), $\boldsymbol{\mu}$ é o vetor imagem média total, c é o número de classes e N_i é o número de amostras da classe Γ_i . A imagem média da classe $\boldsymbol{\mu}_i$ é obtida concatenando-se todas as colunas da imagem para formar um único vetor-coluna (Martínez e Kak, 2001).

Caso a matriz S_W seja não singular, a projeção ótima W_{opt} é escolhida de tal forma que maximize a razão entre o determinante da matriz de variância entre-classes das amostras projetadas e o determinante da matriz de variância intra-classes das amostras projetadas, ou seja,

$$W_{opt} = \arg \max_W \frac{\det(W^T S_B W)}{\det(W^T S_W W)} = [\mathbf{w}_1 \dots \mathbf{w}_m], \quad (3.9)$$

onde $\{\mathbf{w}_i | i = 1, 2, \dots, m\}$ é o conjunto de autovetores de S_B e S_w , correspondendo aos m autovalores $\{\lambda_i | i = 1, 2, \dots, m\}$, assim (Belhumeur et al., 1997),

$$\{S_B \mathbf{w}_i = \lambda_i S_w \mathbf{w}_i\}. \quad (3.10)$$

O espaço de projeções é então encontrado resolvendo a equação (3.8), em que W é a matriz com os autovetores generalizados, \mathbf{w}_i , associados aos autovalores, λ_i , da matriz diagonal. A determinação de um subespaço vetorial W_{opt} que maximize o critério de *Fisher* pode ser resolvida como um problema de autovetores e autovalores.

Uma observação importante a ser feita é em relação ao número de amostras para treinamento, pois caso este seja muito pequeno e a dimensão dos dados seja grande, a inversão da matriz S_w dependerá deste número de amostras que a base de dados possui. Ou seja, com poucas amostras as matrizes a serem estimadas serão singulares, ficando o cálculo da equação (3.9) comprometido ao momento que se avalie o determinante no seu denominador (que poderá tender a zero). Observe que na equação (3.8), para c classes ter-se-á $c-1$ autovalores generalizados diferentes de zero. Como $S_w \in \mathbb{R}^{a \times a}$, e supondo a N_i observação, S_w teria posto igual a $N_i - c$. Se a dimensão a for muito maior que N_i , como acontece regularmente em imagens, S_w se tornaria uma matriz singular e, portanto não inversível. Neste sentido, alguns autores, como (Belhumeur et al., 1997), propuseram métodos para superar esta instabilidade da matriz S_w e sugeriram o uso de um sub-espaço PCA, mas aplicado a métodos de reconhecimento de face, também conhecido como *Fisherfaces*. Para o caso específico deste trabalho, o reconhecimento através da forma de caminhar, este problema não ocorre devido ao grande número de informações contidas nos vetores de características extraídas. Mas, caso ocorresse, teria-se que W_{opt} então seria (Liu e Sarkar, 2006)

$$W_{opt}^T = W_{FLD}^T W_{PCA}^T, \quad (3.11)$$

onde

$$W_{PCA} = \arg \max_W \left| W^T S W \right| \text{ e} \quad (3.12)$$

$$W_{FLD} = \arg \max_W \frac{\left| W^T W_{PCA}^T S_B W_{PCA} W \right|}{\left| W^T W_{PCA}^T S_W W_{PCA} W \right|}. \quad (3.13)$$

W_{PCA} é W^* da seção 3.2 e W_{FLD} é o espaço de projeção de Fisher.

3.4 Proporção por Variâncias

A técnica de Proporção por Variâncias (*Proportion of Variance – POV*) utiliza um método semelhante ao LDA. É calculada a razão entre as variâncias entre-classes e intra-classes das amostras, e as características que possuem maior razão são seleccionadas. Assim, a m -ésima característica do q -ésimo $\{q = 1, \dots, Q\}$ vetor de característica da forma de caminhar do k -ésimo $\{k = 1, \dots, K\}$ indivíduo pode ser expresso como $y_m(k, q)$. A média dos vetores de características para cada classe pode ser obtida por

$$\bar{y}_m(k) = \frac{1}{Q} \sum_{q=1}^Q y_m(k, q), \quad (3.14)$$

e a média dos vetores de características de toda a base de dados é denotado por (Ekenel e Sankur, 2004)

$$\bar{\bar{y}}_m = \frac{1}{KQ} \sum_{k=1}^K \sum_{q=1}^Q y_m(k, q). \quad (3.15)$$

A razão entre a variância entre-classes e a variância intra-classes para a m -ésima coluna do vetor de característica é representada por

$$r_m = \frac{\sum_{k=1}^Q (\bar{y}_m(k) - \bar{\bar{y}}_m)^2}{\sum_{k=1}^K \sum_{q=1}^Q (\bar{y}_m(k, q) - \bar{\bar{y}}_m(k))^2}, \quad (3.16)$$

que é o coeficiente de redução POV.

A equação 3.16 dá a proporção entre estas duas variâncias, para cada característica m . Assim, somente é mantida a característica m se o valor da razão r_m for maior que um limiar. Neste ponto, é realizada uma seleção de características, reduzindo-se a dimensão do vetor de características. Como dito anteriormente, POV se assemelha ao discriminante linear de Fisher (LDA), mas sem sua complexidade de manipular matrizes que podem, inclusive, ser singulares. POV é, portanto, um método de seleção das características mais discriminativas (ao estilo Fisher) do que representativas (como é a técnica PCA). Novamente, PCA pode selecionar características de grande energia, mas que talvez não consiga discriminar bem entre diversas classes. Como POV maximiza a razão entre variância entre classes e a variância dentro da classe, poder-se-ia selecionar aqueles atributos cuja razão seja a maior possível. Mas lembrando que são selecionadas aquelas cuja variância entre as classes tende a ser grande e dentro das classes tende a ser pequena, mesmo que tenham pouca energia em termos representativos.

Capítulo 4

Modelos Ocultos de Markov

4.1 Introdução

Nos capítulos anteriores, foi visto as etapas e técnicas de pré-processamento e redução dimensional para conjuntos de dados de indivíduos que estejam numa base de dados do sistema de reconhecimento de indivíduos pela forma de caminhar proposto, tendo como resultado conjuntos de vetores representativos para cada indivíduo. Neste capítulo, será mostrada a última etapa do sistema proposto, que é a etapa de classificação. Foi utilizado os Modelos Ocultos de Markov (*Hidden Markov Models* – HMM) como classificador de indivíduos, pois esta ferramenta tem a capacidade de captar as características dinâmicas e desta forma diferenciar sequências de vetores de classes distintas (Liu e Sarkar, 2005).

O HMM se enquadra como um modelo estocástico que caracteriza apenas as propriedades estatísticas de um sinal. Este sinal pode ser caracterizado como um processo estocástico, e os parâmetros do processo estocásticos podem ser estimados por uma maneira pré-definida (Rabiner, 1989). Em resumo, o HMM modela uma sequência de observações de um processo estacionário.

Nos últimos anos, os Modelos Ocultos de Markov têm sido intensamente aplicados em várias áreas tais como: padrões, patologias, reconhecimento de fala, análise de sequências de DNA, etc (Bishop, 2006). O HMM é adequado para classificação de sinais de uma ou duas dimensões, e pode ser utilizado também quando a informação é incompleta ou incerta.

Para usar o HMM, é necessário realizar uma fase de treinamento supervisionado, com o ajuste de parâmetros do modelo para, em seguida, realizar-se a fase de teste (classificação, neste trabalho) (Kale et al., 2002). Na fase de treino é utilizado o algoritmo de Baum-Welch para estimação dos parâmetros do modelo, cujo método é baseado no critério da máxima verossimilhança. Já para a fase de teste, computa-se as sequências de estados mais prováveis, geralmente utilizando-se o algoritmo de Viterbi (Rabiner, 1993).

Nas próximas seções serão mostrados, detalhadamente, a teoria básica, os três problemas fundamentais do HMM e a descrição das fases de treino e classificação aplicados ao problema de identificação de indivíduos pela forma do caminhar.

4.2 Modelos Ocultos de Markov

O HMM é um modelo estocástico baseado em equações de estado, sendo os estados a parte oculta do modelo. Os estados são estimados a partir de variáveis observáveis do sistema o qual se deseja modelar. Assim, assume-se uma sequência O constituída de T vetores de observações O_t , $O = \{O_1, O_2, \dots, O_t, \dots, O_T\}$, $t \in \{1, \dots, T\}$. No presente caso, cada vetor O_t representa uma medida sobre as silhuetas do caminhar. A cardinalidade do conjunto $\{1, \dots, T\}$ determina o número de quadros do vídeo capturado. Os parâmetros de interesse de um HMM são (Rabiner, 1993):

- a) o número N de estados no modelo, onde os estados individuais são $S = \{S_i\}$, $i \in \{1, \dots, N\}$, e o estado no tempo t é q_t . Embora os estados sejam ocultos, para muitas aplicações práticas, existe um significado físico associado a estes. Geralmente, os estados estão interconectados de tal maneira que qualquer estado pode se comunicar com outro estado (modelo ergódico). Outros tipos de interconexão entre estados também são possíveis e serão mostrados mais adiante;
- b) o número M de símbolos distintos observados por estado. Os símbolos correspondem à saída do sistema. Denota-se símbolos individuais como v , $V = \{v_1, v_2, \dots, v_M\}$;
- c) o vetor de probabilidade inicial (π), dado por $\pi = \{\pi_i\}$ representa a probabilidade de estar no estado S_i no início do experimento;
- d) a matriz de transição de estados A , formada por entradas a_{ij} que indicam a probabilidade de se estar num estado S_j no tempo $t+1$, dado que o estado anterior era o estado S_i no tempo t . Para o caso especial em que qualquer estado se comunica com qualquer outro estado, tem-se que $a_{ij} > 0$ para todo i, j . Para outros tipos de HMMs, tem-se que $a_{ij} = 0$ para um ou mais (i, j) pares;

- e) o conjunto B das distribuições de probabilidade de se obter na saída a observação no estado S_j . Observe que, neste caso, optou-se por utilizar o modelo contínuo do HMM ao invés do modelo discreto, o que levaria o parâmetro B a ser representado por uma matriz de probabilidades de emissão de símbolos.

A sequência de estados S é gerada por uma máquina de estados probabilística caracterizada pela matriz A e o vetor π . A sequência de observações depende do estado em cada instante e é caracterizada pela matriz de probabilidade B (Marques, 1999) (Liu e Sarkar, 2005).

Convenciona-se representar cada modelo λ pela tripla (π, A, B) (Kale et al., 2004). Têm-se então as seguintes equações:

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i), \quad (4.1)$$

$$b_j(o_t) = P(o_t | q_t = S_j), \quad (4.2)$$

$$\pi_i = P(q_1 = S_i). \quad (4.3)$$

Tendo-se os valores apropriados de N , M e $\lambda = (\pi, A, B)$, o HMM pode ser utilizado como um gerador de sequência de observações. No caso do presente trabalho, estas observações já são conhecidas.

O HMM pode ser classificado de dois tipos principais: o modelo ergódico (cheio) ou o modelo esquerda-direita, como representado nas Figuras 4.1 (a) e (b). O modelo ergódico é o modelo onde o HMM é totalmente conectado, ou seja, todos os estados são conectados entre si. Desta forma, a matriz A é uma matriz cheia, daí a sua denominação. O modelo esquerda-direita é aquele que conforme o tempo é incrementado, o índice do estado também é incrementado, resultando num sentido de sequências de estados “esquerda-direita”. Este modelo tem como propriedade $a_{ij} = 0$, para todo $j < i$.

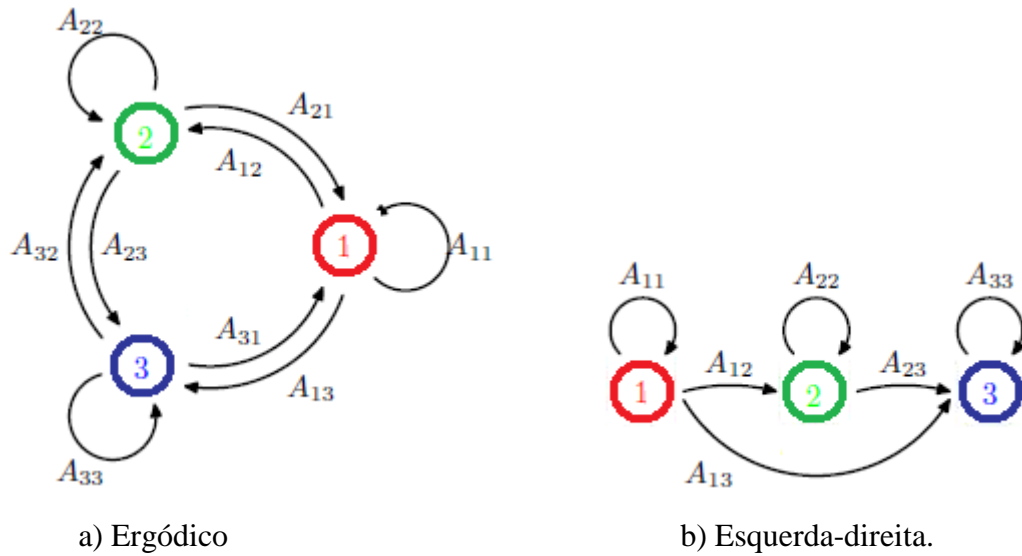


Figura 4.1: Dois exemplos de topologias para HMM's. O modelo (b) foi utilizado neste trabalho.

4.3 Os Três Problemas Básicos do HMM

Existem três problemas clássicos a serem solucionados para o modelo HMM. Os problemas são os seguintes:

Problema 1: Dada a sequência de observações $O = \{O_1, O_2, \dots, O_t, \dots, O_T\}$ e o modelo $\lambda = (\pi, A, B)$, qual a probabilidade da sequência de observações ser produzida pelo modelo, $P(O|\lambda)$?

Problema 2: Dada a sequência de observações $O = \{O_1, O_2, \dots, O_t, \dots, O_T\}$ e o modelo $\lambda = (\pi, A, B)$, qual a sequência de estados S que melhor representa as observações?

Problema 3: Como se ajusta os parâmetros do modelo $\lambda = (\pi, A, B)$ para maximizar $P(O|\lambda)$?

O Problema 1 pode ser visto como quão bom um dado modelo representa uma sequência de observações. Este problema se aplica à etapa de classificação (ou teste) de um sistema de reconhecimento de padrões. O Problema 2 tenta descobrir qual sequência de estados melhor se adequa a um dado conjunto de observações. Vale ressaltar que não existe a

correta sequência de estados para uma dada sequência de observações, mas sim aquela sequência de estados que melhor representa as observações, tendo em vista que se faz o uso de um critério ótimo para solucionar este problema. O Problema 3 ajusta os parâmetros (π, A, B) , para melhor representar uma sequência de observações. A sequência de observações utilizada para ajustar os parâmetros é chamada de conjunto de treino, pois é ela que vai treinar o HMM. Esta é uma etapa muito importante pois com o ajuste adequado dos parâmetros criam-se modelos que melhor representam um dado conjunto de observações reais.

As três subseções a seguir irão mostrar detalhadamente como são obtidas as soluções dos três problemas do HMM já apresentados.

4.3.1 Solução do Problema 1

Pretende-se calcular a probabilidade da sequência de observação, $O = \{\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_t, \dots, \mathbf{O}_T\}$, dado o modelo $\lambda = (\pi, A, B)$, isto é, $P(O|\lambda)$. A maneira mais simples de se obter este resultado é enumerando cada sequência de estado. Considere uma sequência de estados fixa $Q = \{q_1, q_2, \dots, q_T\}$, onde q_1 é o estado inicial. A probabilidade da sequência O para a sequência de estados Q é (Rabiner, 1993)

$$P(O|Q, \lambda) = \prod_{t=1}^T P(\mathbf{O}_t | q_t, \lambda), \quad (4.4)$$

onde se assume independência entre as observações. Assim,

$$P(O|Q, \lambda) = b_{q_1}(\mathbf{O}_1) \cdot b_{q_2}(\mathbf{O}_2) \cdot \dots \cdot b_{q_T}(\mathbf{O}_T). \quad (4.5)$$

A probabilidade de cada sequência de estado Q pode ser escrita como

$$P(Q|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \dots a_{q_{T-1} q_T}. \quad (4.6)$$

A probabilidade conjunta de O e Q é

$$P(O, Q|\lambda) = P(O|Q, \lambda) P(Q, \lambda) \quad (4.7)$$

A probabilidade de O dado o modelo é a soma de todas as probabilidades conjuntas para todas as possibilidades de sequências de estados

$$P(O|\lambda) = \sum_{all Q} P(O|Q, \lambda) P(Q, \lambda) = \sum_{q_1, \dots, q_T} \pi_{q_1} b_{q_1}(\mathbf{O}_1) a_{q_1 q_2} b_{q_2}(\mathbf{O}_2) \dots a_{q_{T-1} q_T} b_{q_T}(\mathbf{O}_T). \quad (4.8)$$

A equação (4.8) não pode ser usada na prática, pois envolve um número de operações inportável ($2TN^T$ operações) (Rabiner, 1993). É possível ultrapassar esta dificuldade usando os algoritmos *progressivo* e *regressivo* (algoritmo *forward* e *backward*), desenvolvido por Baum (Baum, 1972), onde o custo computacional fica na ordem de N^2T . O primeiro é baseado na *probabilidade progressiva*, ou seja, na probabilidade da sequência parcial das observações até o tempo t , dado o modelo, definida como

$$\alpha_t(i) = P(\mathbf{O}_t, q_t = S_i | \lambda). \quad (4.9)$$

Pode-se resolver (4.9) intuitivamente, considerando que

$$\alpha_1(i) = \pi_i b_i(\mathbf{O}_1), \quad 1 \leq i \leq N. \quad (4.10)$$

Desta forma, tem-se que

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(\mathbf{O}_{t+1}), \quad 1 \leq t \leq T-1, \quad 1 \leq j \leq N, \quad (4.11)$$

e chegando a

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i). \quad (4.12)$$

O mesmo se aplica à probabilidade regressiva,

$$\beta_t(i) = P(\mathbf{O}_{t+1} | q_t = S_i, \lambda). \quad (4.13)$$

Intuitivamente tem-se que

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad (4.14)$$

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, \dots, 1, \quad 1 \leq i \leq N. \quad (4.15)$$

Desta forma, a probabilidade da sequência observada é

$$P(O|\lambda) = \sum_{j=1}^N \pi_j b_j(O_1) \beta_1(j). \quad (4.16)$$

4.3.2 Solução do Problema 2

Diferentemente do problema 1, em que é possível obter uma solução exata, no problema 2 tenta-se achar uma solução ótima, ou seja, uma sequência de estados ótima, que melhor represente uma dada sequência de observação.

Há duas formas de solucionar este problema. A primeira é fazendo uma abordagem onde se escolhem os estados q_t que melhor representem individualmente a sequência de observações. A segunda é fazendo uma abordagem em que modifica o critério de otimização. Assim, para a primeira abordagem, define-se a variável

$$\gamma_t(i) = P(q_t = S_i | O, \lambda), \quad (4.17)$$

que é a probabilidade de estar no estado S_i dada uma sequência de observação O e modelo λ . Esta mesma equação também pode ser descrita como

$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i) \beta_t(i)}{\sum_{i=1}^N \alpha_t(i) \beta_t(i)}, \quad (4.18)$$

para $\alpha_t(i)$ tratando das observações O_1, O_2, \dots, O_t e com o estado S_i no tempo t , e $\beta_t(i)$ tratando das observações $O_{t+1}, O_{t+2}, \dots, O_T$, dado o estado S_i no tempo t . Observe que o denominador, $P(O|\lambda)$, faz com que

$$\sum_{i=1}^N \gamma_t(i) = 1. \quad (4.19)$$

Utilizando (4.17), pode-se concluir que o melhor estado individual é

$$q_t = \arg \max_{1 \leq i \leq N} [\gamma_t(i)], \quad 1 \leq t \leq T. \quad (4.20)$$

Apesar desta solução determinar qual o melhor estado para cada tempo t , nos casos em que $a_{ij} = 0$, o estado em questão não representa a melhor solução. Uma forma de se obter uma sequência de estados ótima é modificando o critério de otimização. Ou seja, pode-se escolher as melhores sequências de estados aos pares, ou em trios, e assim por diante. Embora este critério se torne mais razoável para algumas aplicações, o que mais se utiliza é descobrir uma única melhor sequência de estados, ou seja, que maximize $P(Q|O, \lambda)$, que é o equivalente a maximizar $P(Q, O|\lambda)$. A forma na qual é possível se obter tal sequência é através de métodos de programação dinâmica, mais conhecida como algoritmo de Viterbi (Marques, 1999).

O algoritmo de Viterbi foi originalmente desenvolvido para a decodificação de códigos convolucionais, especialmente utilizados em sistemas de comunicação de dados. O algoritmo foi criado de modo a maximizar a verossimilhança de uma sequência de informação decodificada. O algoritmo de Viterbi pode ser interpretado como uma aplicação do Princípio da Otimalidade de Bellman (Bellman, 1957), que diz que um método de otimização global pode ser obtido por uma sequência ótima de processos localmente otimizados. O uso do algoritmo de Viterbi também se popularizou na determinação da máxima verossimilhança para modelos de Markov, com o estudo de modelos estocásticos, onde se infere um processo oculto (latente) a partir de um processo observável.

A sequência de estados mais provável pode ser calculada, definindo-se uma variável $\delta_t(i)$ expressa por

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(Q, O|\lambda). \quad (4.21)$$

Por indução tem-se

$$\delta_{t+1}(j) = [\max_i \delta_t(i) a_{ij}] \cdot b_j(O_{t+1}). \quad (4.22)$$

A sequência ótima é obtida por retrossuigmento através das seguintes equações:

$$p^* = \max_{1 \leq i \leq N} [\delta_T(i)], \quad (4.23)$$

$$q_t^* = \arg \max_{1 \leq i \leq N} [\delta_t(i) a_{ip^*_{t+1}}], \quad t = T-1, T-2, \dots, 1. \quad (4.24)$$

4.3.3 Solução do Problema 3

Este último problema é considerado o mais difícil de ser resolvido. Este problema requer que se determine um método para ajuste dos parâmetros do modelo, (π, A, B) , de forma a maximizar a probabilidade de uma sequência de observações. A dificuldade está em se estimar um modelo oculto A (variável latente, portanto não observável) a partir das observações. O modelo A constitui-se de uma representação no espaço de estados com transições estocásticas, onde as probabilidades de permanência ou salto de um estado para o outro devem ser inferidas.

No que tange à maximizar a probabilidade de uma sequência de observações, o maior obstáculo está no fato de que não existe uma solução analítica que maximize esta probabilidade. Entretanto, é possível escolher os parâmetros do modelo de forma a que se maximize localmente esta probabilidade através de um método iterativo chamado de algoritmo de re-estimação de Baum-Welch. O algoritmo de re-estimação de Baum-Welch é equivalente ao método da Máxima Verossimilhança e emprega a técnica conhecida como *Expectation and Maximization* - EM (Rabiner, 1993).

A fim de descrever os procedimentos para a re-estimação dos parâmetros, primeiro considera-se que $\xi_t(i, j)$ seja a probabilidade de se estar no estado S_i no tempo t , e de se estar no estado S_j no tempo $t+1$, dado o modelo e a sequência de observações, ou seja,

$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda), \quad (4.25)$$

$$= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O | \lambda)}, \quad (4.26)$$

$$= \frac{\alpha_t(i) a_{ij} b_j(\mathbf{O}_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(\mathbf{O}_{t+1}) \beta_{t+1}(j)}. \quad (4.27)$$

Da equação (4.17), pode-se relacionar $\gamma_t(i)$ com $\xi_t(i, j)$, para todo j , por

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j). \quad (4.28)$$

Desta forma, pode-se dizer que $\sum_{t=1}^{T-1} \gamma_t(i)$ é o número de transições do S_i e

$\sum_{t=1}^{T-1} \xi_t(i, j)$ é o número de transições de S_i para S_j .

Baseado neste tipo de contagem de ocorrências, será apresentado um método de re-estimação dos parâmetros de um HMM, cujas fórmulas de re-estimação são dadas por:

$$\bar{\pi}_i = \text{frequência esperada no estado } S_i \text{ no tempo } (t=1) = \gamma_1(i), \quad (4.29)$$

$$\bar{a}_{ij} = \frac{\text{número esperado de transições do estado } S_i \text{ para o estado } S_j}{\text{número esperado de transições do estado } S_i} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, \quad (4.30)$$

$$\bar{b}_j(k) = \frac{\text{número esperado de vezes no estado } j \text{ observando } v_k}{\text{número esperado de vezes no estado } j} = \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \quad (4.31)$$

Se iterativamente for utilizado $\bar{\lambda}$ no lugar de λ , e repetir-se a re-estimação, chega-se a um limite e o resultado do procedimento de re-estimação é chamado de algoritmo *de re-estimação de Baum-Welch*, que é um caso particular do algoritmo EM. O algoritmo de Baum-Welch consiste nos seguintes passos (Rabiner, 1993):

Passo 1) Inicialização: Escolha dos parâmetros iniciais do modelo. Pode-se fazer de forma aleatória ou utilizando algum tipo de estimativa para inicialização.

Passo 2) Passo E (Estimação): Cálculo das esperanças matemáticas das variáveis ocultas, dado λ , ou seja, o número de vezes nos estados, transições dos estados e outras variáveis indicadas no **Passo M**.

Passo 3) Passo M (Maximização): Cálculo dos novos parâmetros $\bar{\lambda}$ através de (4.29), (4.30) e (4.31) para que $P(O|\bar{\lambda})$ seja maximizado utilizando os valores calculados no **Passo E**.

Passo 4) Iteração: Se $P(O|\bar{\lambda}) - P(O|\lambda)$ for maior que um valor fixo, ir para o **Passo E**. Caso contrário, terminar a execução e retornar os novos parâmetros.

Foi provado por Baum e seus colegas que $P(O|\bar{\lambda}) \geq P(O|\lambda)$ e que a sequência de modelos $\bar{\lambda}_i$ obtidos com o EM converge para λ^* , um máximo local de $\arg \max_{\lambda} P(O|\lambda)$. O algoritmo de Baum-Welch converge localmente para um máximo razoável e é computacionalmente rápido (Rabiner, 1993).

4.4 O HMM Aplicado à Dinâmica do Caminhar

Para um sistema robusto de reconhecimento de indivíduos pela forma de caminhar é razoável que o vetor de características seja formado por um conjunto de ciclos, e não por apenas um ciclo (Kale et al., 2002). Desta forma, o problema em questão é reconhecer um indivíduo que está num banco de dados cujo modelo de caminhar já se conhece a priori.

Na etapa de treino, estimam-se os parâmetros do modelo empregando-se o algoritmo de Baum-Welch, segundo a seguinte operação:

$$\arg \max_{\lambda_i} P(O_i | w_i, \lambda_i), \quad (4.32)$$

onde λ_i se refere ao modelo associado à i-ésima classe w_i (i-ésimo indivíduo).

Para a etapa de classificação, a partir de uma sequência de caminhar O , calcula-se qual dos modelos se aproxima mais dos dados a serem testados. Inicialmente, a sequência de imagens $O=\{O_1, O_2, \dots, O_t, \dots, O_T\}$ do indivíduo desconhecido passa pelo mesmo processamento que as imagens utilizadas como treino no sistema. Assim, basta computar a verossimilhança que a sequência de observações gerada a partir do HMM correspondente à i -ésima pessoa. Este procedimento é aplicado a cada pessoa do banco de dados e vai resultar num valor de probabilidade (Vieira, et al., 2009). A classificação ou identificação do indivíduo desconhecido é obtida quando este valor de probabilidade resulta no maior valor entre todas as classes existentes, ou seja, a classe i^* que obtiver o maior valor será a classe correspondente para a sequência de imagens do indivíduo desconhecido em questão (Kale et al., 2002) (Nikolaos et al., 2005). Portanto, o processo de identificação realiza os seguintes passos:

Passo 1) Avalia-se, para cada modelo λ_i , a melhor sequência de estados Q^* que explica O ,

Passo 2) Para cada Q^* (que representa um determinado modelo) obtém-se a probabilidade de emissão de O ,

Passo 3) Classifica-se O associando-o ao modelo indexado por i^* que apresentar a sequência de estados mais provável.

Estes passos podem ser sintetizados na expressão:

$$i^* = \underset{i}{\operatorname{argmax}} P(O|w_i, \lambda_i, Q_i^*), i \in \{1, \dots, M\}. \quad (4.33)$$

Capítulo 5

Procedimentos Experimentais e Resultados

5.1 Metodologia e Dados Utilizados

O sistema de identificação de indivíduos pela forma de caminhar proposto já pode ser apresentado de forma mais detalhada, tendo em vista que todas as etapas de pré-processamento, as técnicas de redução dimensional e o tipo de classificador já foram apresentados e abordados com detalhamento teórico nos capítulos anteriores. Neste capítulo de procedimentos experimentais e resultados serão especificados as bases de dados utilizadas, os valores dos parâmetros empregados no algoritmo implementado, as limitações dimensionais encontradas e as soluções tomadas visando um resultado mais eficiente, e, sobretudo, apresentar os resultados obtidos através de uma análise crítica.

As bases de dados utilizadas foram duas, uma com um grande número de indivíduos, 100 pessoas, e a outra com um número um pouco reduzido, 16 pessoas, com a finalidade de avaliar o comportamento do sistema de reconhecimento para conjuntos populacionais grandes e pequenos. A base de dados com 100 pessoas é a base CASIA (conjunto B) (CASIA), uma base de dados disponibilizada pelo Instituto de Automação da Academia Chinesa de Ciências (*Institute of Automation, Chinese Academy of Sciences*), que está no formato AVI com razão de captura de 25 quadros por segundo, 6 sequências por indivíduo (3 para treino e 3 para teste). Esta base de dados é composta por 11 ângulos de visão, mas foram utilizados apenas os vídeos cujo ângulo do indivíduo em relação à câmera é de 90° (vista de perfil do indivíduo). Foram utilizados vídeos com a pessoa caminhando normalmente, com a pessoa vestindo roupas diferenciadas (roupas mais largas, como casacos) e com a pessoa usando algum tipo de bolsa. O uso destes vídeos com situações diversas tem por finalidade verificar o comportamento do sistema, e se este é robusto frente a diferentes situações a qual o indivíduo pode estar no momento da captura do vídeo (Yu et al., 2006). A base de dados com 16 pessoas é a base GATECH (GATECH), da Faculdade de Computação do Instituto de

Tecnologia da Geórgia (*Georgia Institute of Technology*), também em formato AVI e razão de captura de 33 quadros por segundo, 5 sequências por indivíduo (3 para treino e 2 para teste).

A sequências de vídeos de ambas as bases de dados possuem aproximadamente de 65 a 120 quadros cuja dimensão é de 320×240 pixels. As sequências possuem imagens de indivíduos caminhando próximo e longe da câmera, coloridas, mas o atributo cor não foi utilizado, e portanto os vídeos foram convertidos do formato RGB para nível de cinza.

Conforme apresentado no Capítulo 3 , foram testados três tipos de técnicas de redução dimensional. Portanto, os parâmetros envolvidos e resultados obtidos de identificação dos indivíduos serão apresentados de acordo com estas três técnicas, com a finalidade de se fazer uma comparação entre os métodos usados ao final da apresentação de cada uma. Para os resultados referentes à utilização do banco de dados CASIA, com os indivíduos caminhando normalmente, será feita referência a estas sequências de imagens como CASIA A e para os indivíduos caminhando vestidos e usando bolsa será feita sua referência como CASIA B, como mostrado na Figura 5.1. Por outro lado para as sequências em que os indivíduos fazem parte da base de dados GATECH, apenas será mencionado GATECH (ver Figura 5.2).

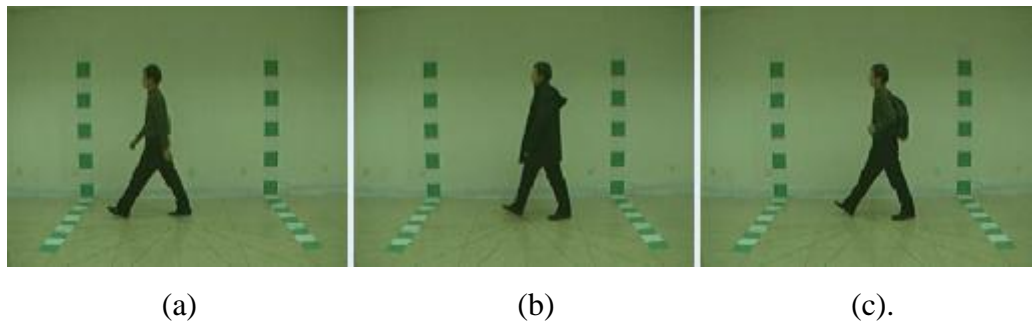


Figura 5.1: Imagens das bases de dados CASIA A (a) e B (b) e (c).



Figura 5.2: Imagem da base de dados GATECH.

5.2 Método de Redução Dimensional Usando PCA

De posse das sequências de vídeo das bases de dados mencionadas, fez-se um pré-processamento, como descrito no Capítulo 2, para a obtenção do vetor de característica da forma de caminhar de cada pessoa. Como cada indivíduo possui altura diferente e a posição do indivíduo em relação à câmera é variável, a dimensão do vetor varia tanto em função da característica física da pessoa como da distância a que esta caminha em relação à câmera de captura. Deste modo, entre as M linhas que formam a medida da silhueta, fez-se uma redução dimensional, passando o vetor para quarenta linhas, tendo em vista que cada indivíduo tinha entre 50 e 100 linhas. A sequência de observações do vetor de características passa a ter, então, dimensão $40 \times N$, onde N é o número de quadros de cada sequência de imagens. Não é necessário que o número de quadros N seja igual para cada pessoa, pois cada indivíduo possui uma dinâmica do caminhar que o identifica, uns caminham mais lentamente, outros mais rapidamente. Desta forma, não se limitou o parâmetro N para que não se perdesse informações que poderiam ser relevantes. Esta diferença entre pessoas é absorvida no HMM através da transição de estados.

Após a etapa de pré-processamento, o vetor de característica é reduzido em sua dimensão através da técnica do PCA, afim de que as informações a serem utilizadas no classificador HMM possam ser suficientes para um bom desempenho do sistema.

Nas Figuras 5.3, 5.4 e 5.5, estão ilustrados os gráficos que relacionam a variância acumulada em função do número de componentes principais de um indivíduo das duas bases de dados. O número de componentes principais total é 40, pois foram formados vetores de dimensão $40 \times N$ para cada pessoa, e a técnica PCA vai reduzir exatamente este número de linhas do vetor, para que se utilize apenas aquelas componentes que sejam mais importantes.

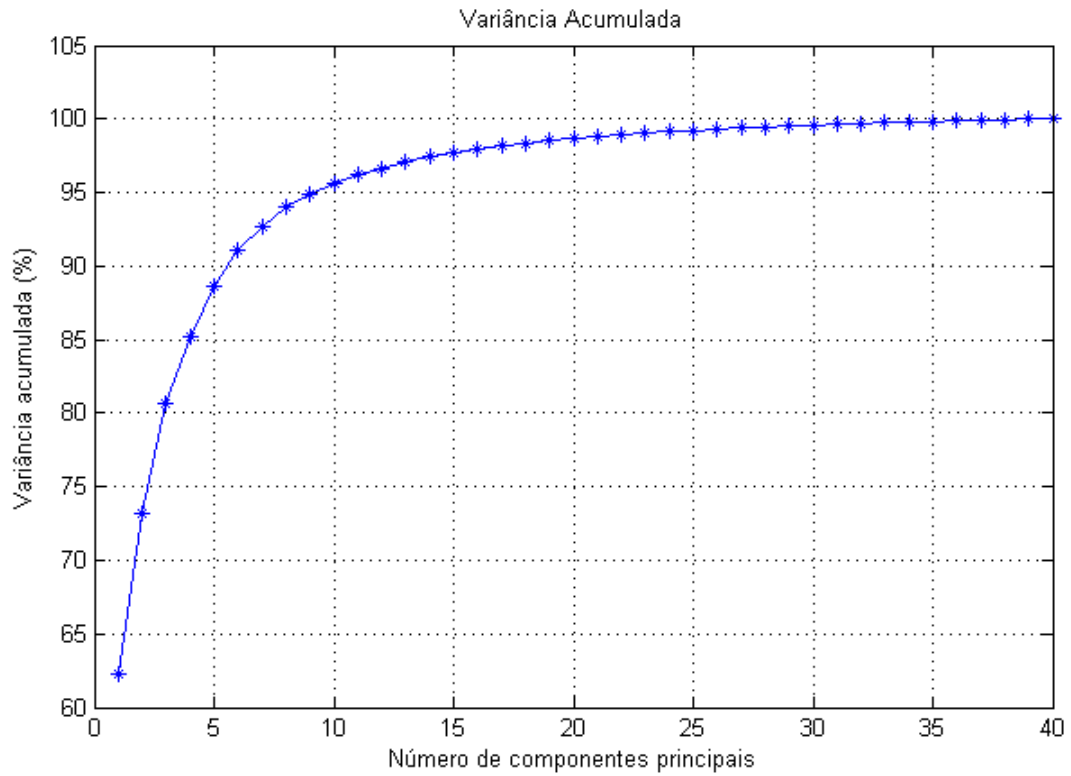


Figura 5.3: Gráfico da variância acumulada em função do número de componentes principais para a base de dados CASIA A.

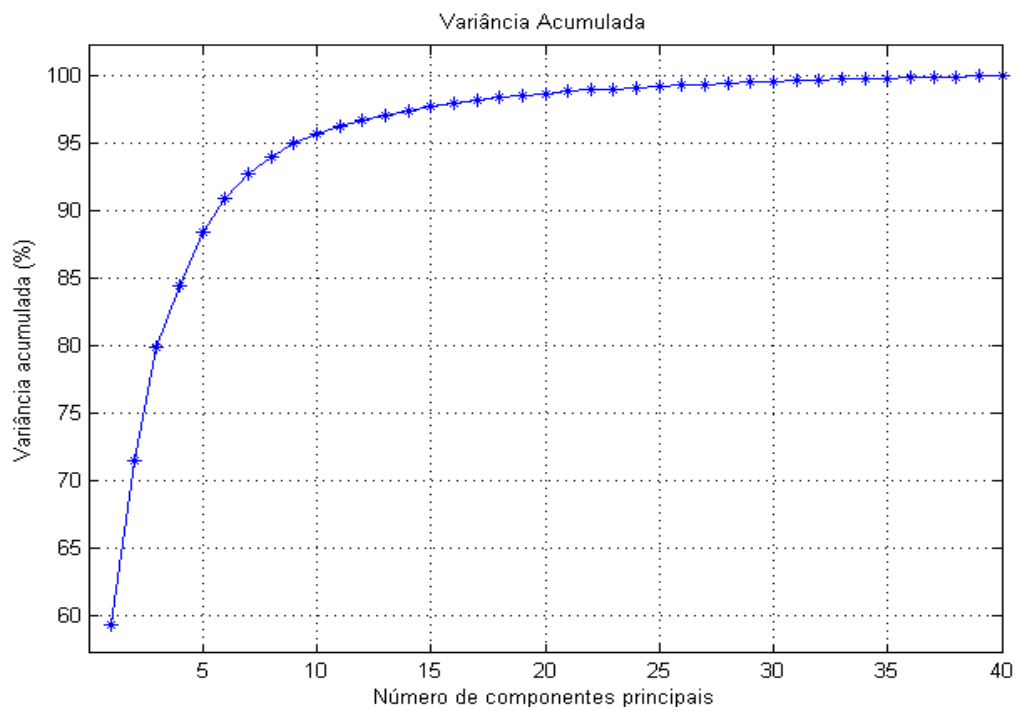


Figura 5.4: Gráfico da variância acumulada em função do número de componentes principais para a base de dados CASIA B.

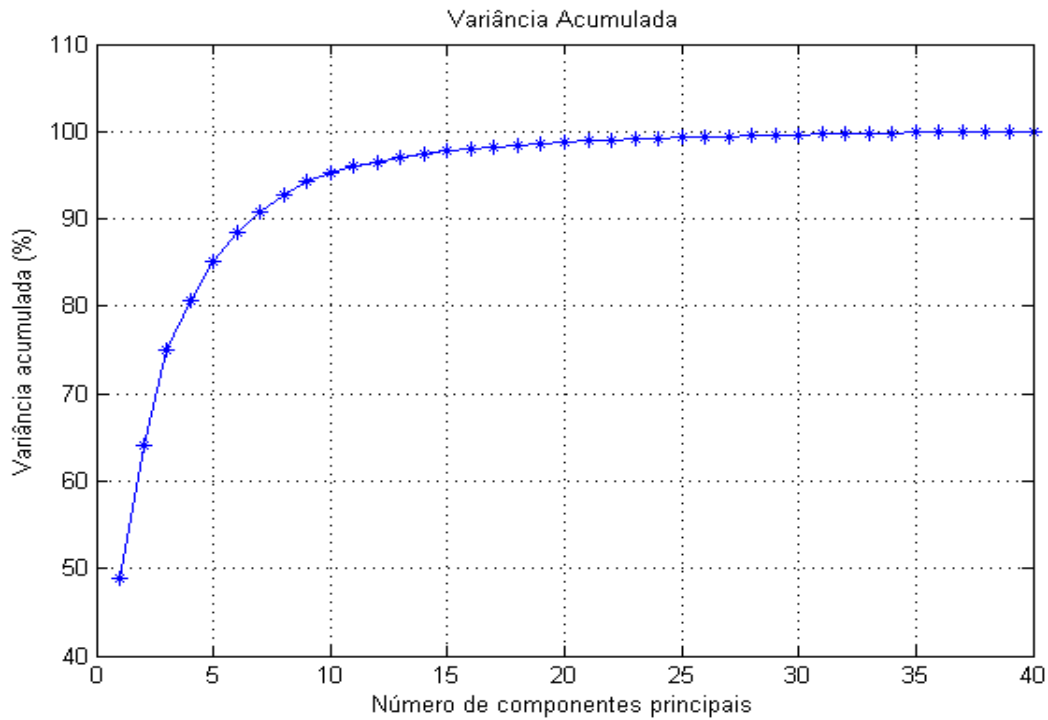


Figura 5.5: Gráfico da variância acumulada em função do número de componentes principais para a base de dados GATECH.

É possível observar o número de componentes principais capaz de explicar uma determinada porcentagem da variância total observada no conjunto de dados da forma de caminhar. Assim, optou-se por trabalhar com um número de componentes que explique pelo menos 90% da variância total, de forma a se ter um conjunto de dados que seja o mais representativo para um determinado indivíduo. Para as figuras 5.3 e 5.4, o número mínimo são de 6 componentes e para a Figura 5.5 são de 7 componentes. Provavelmente, na etapa de classificação dos indivíduos, os números de componentes que poderão resultar em taxas de acerto de classificação mais satisfatórias serão acima dos mencionados, pois só acima destes valores é que se poderá ter uma melhor representatividade.

Considerando que os dados já foram reduzidos em sua dimensão pelo PCA, foi realizada a classificação através do HMM. O HMM utilizado é do tipo esquerda-direita cujo número de estados utilizados foram 6, caracterizando o ciclo de caminhar de um indivíduo (como mostrado nas Figuras 2.6, 2.7 e 2.8) e utilizando apenas uma Gaussiana, já que o modelo é contínuo.

Primeiramente, fez-se o treinamento do HMM onde são estimados os parâmetros dos modelos de cada classe de pessoas com os vetores de treino. A seguir, já com o vetor de teste

do indivíduo desconhecido, foi calculada a probabilidade para que o algoritmo identifique qual a sua classe correspondente.

Os vetores utilizados na fase de treinamento são formados por três sequências de vídeo e para classificação duas e três sequências para as bases GATECH e CASIA, respectivamente. Estas sequências de vídeo foram permutadas entre si, ou seja, as sequências utilizadas no treino posteriormente foram também utilizadas na classificação, com a finalidade de se obter um resultado mais fiel. O resultado final é uma média dos resultados individuais de cada permutação do conjunto de vetores.

As Tabelas 5.1, 5.2 e 5.3 abaixo mostram os resultados para o número de componentes que tiveram taxas de acerto de classificação mais expressivas.

Tabela 5.1: Taxa de acerto (%) pela abordagem PCA para a base de dados CASIA A.

Número de Componentes principais	12	14	19	24	26
Taxa de acerto (%)	77	75	84	82	79

Tabela 5.2: Taxa de acerto (%) pela abordagem PCA para a base de dados CASIA B.

Número de Componentes principais	18	20	24	25	32
Taxa de acerto (%)	73	76	80	78	71

Tabela 5.3: Taxa de acerto (%) pela abordagem PCA para a base de dados GATECH.

Número de Componentes principais	7	8	9	10	15
Taxa de acerto (%)	62,5	81,25	75	68,75	68,75

Fazendo uma análise das taxas de acertos, para cada uma das bases de dados, vê-se que a base CASIA A obteve a maior taxa de acerto, 84%, para 19 componentes principais, a

base CASIA B obteve a maior taxa, 80%, para 24 componentes principais e a base GATECH obteve a maior taxa, 81,25%, para 8 componentes principais. Como o banco de dados GATECH possui um número reduzido de indivíduos (16 pessoas), o classificador necessitou de um número de componentes menor para distinguir estes poucos indivíduos do banco de dados. Porém, para os bancos de dados CASIA A e CASIA B, que possuem 100 indivíduos, foi necessário um número maior de componentes. Este fato ocorre, provavelmente, porque a distinção entre indivíduos seja mais complicada uma vez que se tem mais variabilidade, necessitando de mais informações para que haja uma melhor distinção entre os indivíduos.

Vale ressaltar que o método PCA transforma os vetores de características para uma nova base de forma que cada classe tenha uma melhor representatividade. A nova base decompõe o sinal de modo a se ter associado a cada componente (autovetor) da base um grau de importância (autovalor) daquela componente. Entretanto, isto não significa que se mantivermos aquelas componentes associadas aos maiores graus de importância, teremos uma boa distinção entre classes.

Em relação à redução dimensional da PCA, por exemplo, para a base GATECH o vetor de dados passa a ter dimensão $8 \times N$, o que torna o sistema de reconhecimento mais rápido na fase de classificação, diminuindo o tempo de processamento.

5.3 Método de Redução Dimensional Usando LDA

A partir dos resultados apresentados através do método de redução dimensional usando PCA, na seção 5.2, percebeu-se que o conjunto de dados já previamente simplificado pelo PCA teve uma boa representatividade. Entretanto, esta representatividade não garante que estes dados sejam distintos entre as diversas classes de indivíduos. Desta forma, buscou-se um método que tivesse não só a capacidade de reduzir a dimensão dos dados, mas também de transformar estas informações representativas de cada classe de forma que estes se diferenciem em relação às classes existentes. Verificou-se que a LDA tem esta característica, e se fez o uso desta ferramenta para avaliar seu comportamento para imagens da forma de caminhar.

Os vetores de características para cada classe de indivíduos possuem dimensão $40 \times N$, mostrando que também não se limitou o número de amostras para cada indivíduo. Mas por

questões de dimensionamento do algoritmo utilizado, o vetor foi redimensionado depois para 40×100 . As taxas de acerto para as 3 bases de dados usadas estão ilustradas na Figura 5.6.

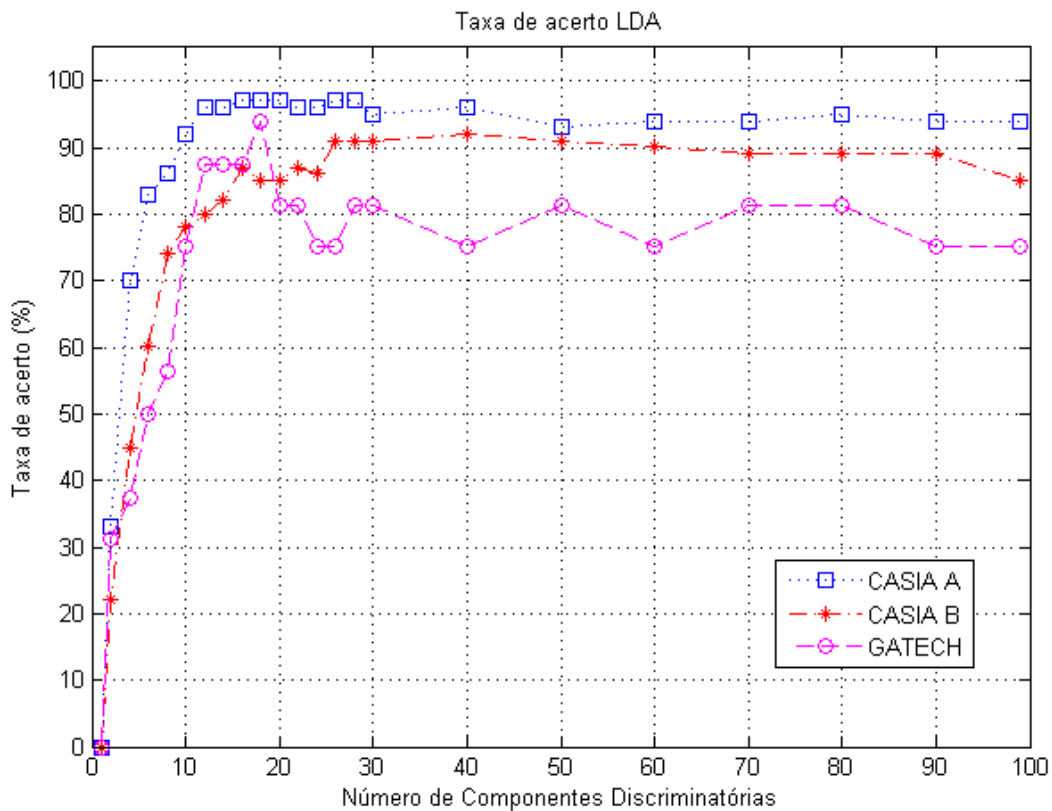


Figura 5.6: Gráfico da taxa de acerto pelo método LDA em função do número de componentes discriminatórias.

A partir dos resultados mostrados, observa-se que o comportamento do HMM frente às 3 bases de dados é semelhante, havendo uma tendência de crescimento rápido, em seguida atinge-se uma faixa de alta taxa de acerto, e depois tende-se a uma estabilização de acertos. Uma faixa de número de componentes discriminatórias é característica para cada base.

Para a base CASIA A, a partir de 10 componentes já se têm acertos acima de 90%, chegando a atingir até 97% com apenas 16 componentes. Na base CASIA B, esta faixa é acima de 12 componentes, não muito diferente da base CASIA A, mas atingiu valores inferiores de acertos. Seu máximo foi de 92%, mas já para um número alto de componentes (40 componentes). Na base GATECH, a taxa de acertos máxima foi de 93,75% e ocorreu com 18 componentes discriminatórias, mas sua faixa de acertos se restringiu ao uso de 12 a 22 componentes.

Comparando o LDA com o PCA, vê-se que os resultados para o LDA, em geral, foram bem superiores que os do PCA, até mesmo se considerarmos o número de componentes (principais e discriminatórias). No PCA houve uma limitação da taxa de acerto para um valor de componentes principais, no LDA houve uma faixa de componentes discriminatórias que garantia altas taxas de acertos. A partir destas observações, conclui-se que o LDA possui um poder de discriminação entre classes muito grande, o que facilitou a classificação realizada pelo HMM.

Para complementar os resultados obtidos quando se tem um grande número de amostras, considerou-se o caso quando se têm poucas amostras, e se fez o teste usando PCA-LDA. Nesta situação, os conjuntos de dados continuaram a ter dimensão $40 \times N$, mas para se utilizar a PCA-LDA é necessário que todos os vetores tenham mesmo tamanho. Assim, os vetores foram redimensionados para 40×100 por uma técnica de interpolação bilinear (técnica do “vizinho mais próximo”) (Gonzalez e Woods, 2002). Como são utilizadas 3 sequências de vídeo para treinamento do HMM, os dados foram colocados em sucessão de modo a formar um vetor-coluna para cada sequência de vídeo. Formou-se 3 vetores-coluna, cada uma de tamanho 4000 para cada classe de pessoas. Em seguida, foi aplicada a técnica PCA, reduzindo os vetores de características para aquelas dimensões que melhor tiveram desempenho. Será chamada de dpca a dimensão para o qual o vetor foi reduzido.

Depois de aplicada a PCA, foi usada a LDA com a finalidade também de redução dimensional, mas de forma a se ter uma diferenciação entre-classes. Como parâmetros para simplificação dos dados foram utilizadas as componentes principais que sejam mais discriminativas. As Figuras 5.7, 5.8, 5.9, 5.10, 5.11 e 5.12 ilustram os resultados de classificação através da taxa de acerto em função da dpca e do número de componentes mais discriminatórias.

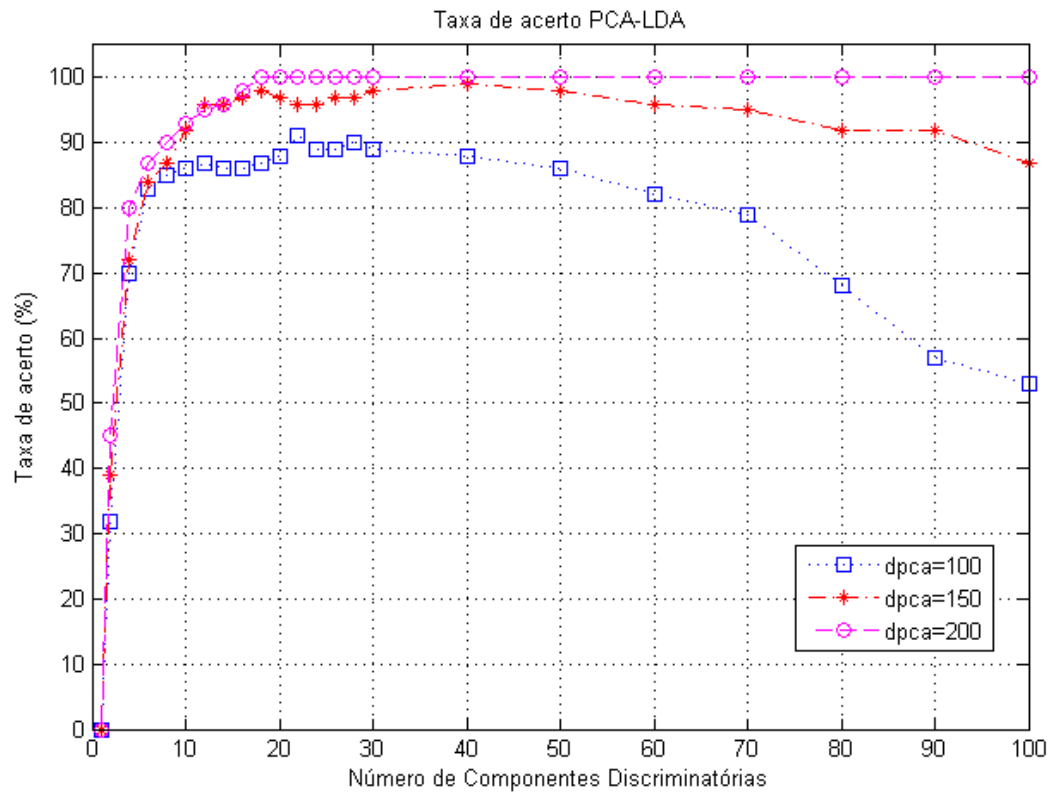


Figura 5.7: Gráfico da taxa de acerto pelo método PCA-LDA para dpca igual a 100, 150 e 200, base de dados CASIA A.

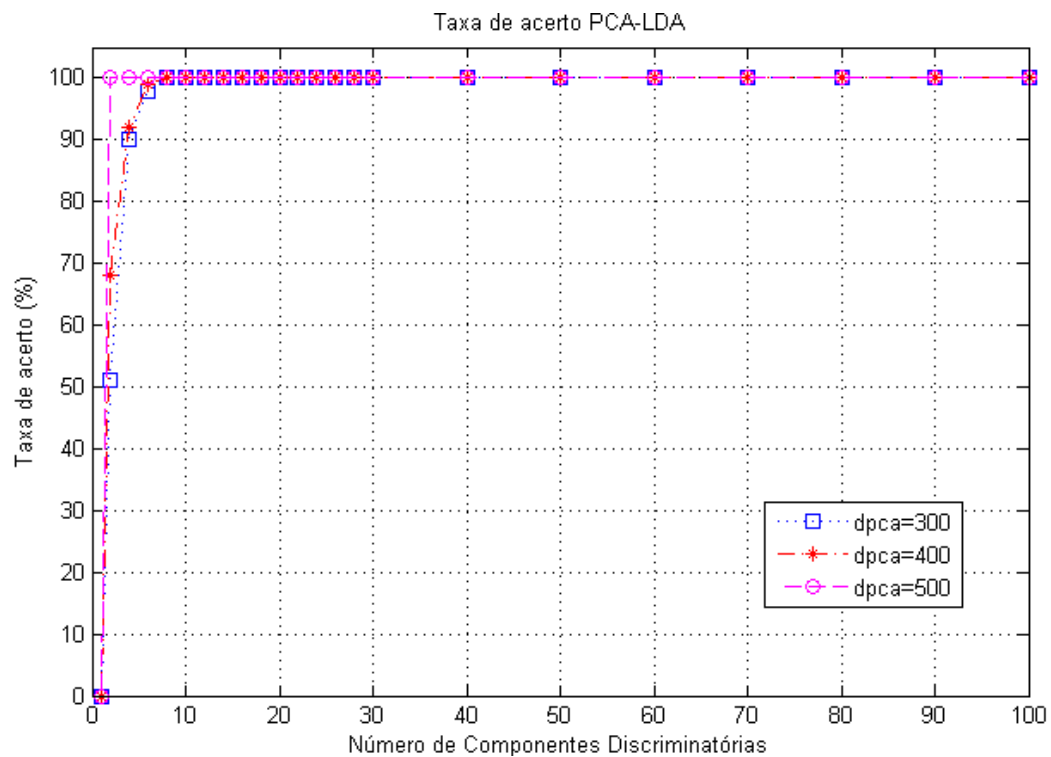


Figura 5.8: Gráfico da taxa de acerto pelo método PCA-LDA para dpca igual a 300, 400 e 500, base de dados CASIA A.

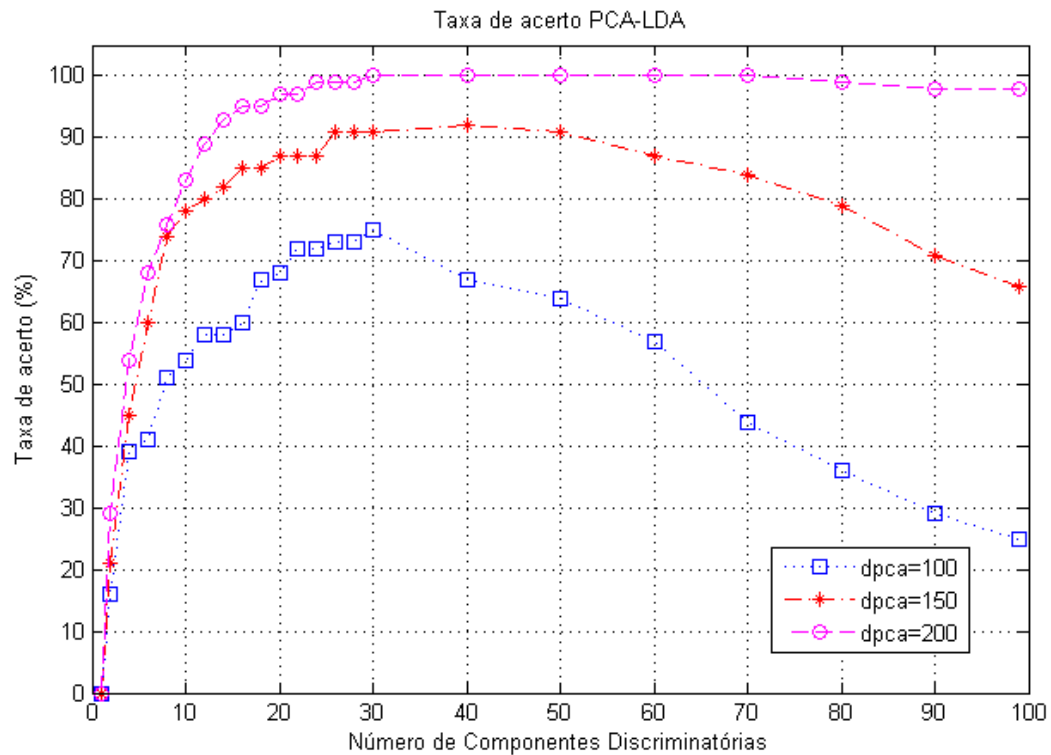


Figura 5.9: Gráfico da taxa de acerto pelo método PCA-LDA para dpca igual a 100, 150 e 200, base de dados CASIA B.

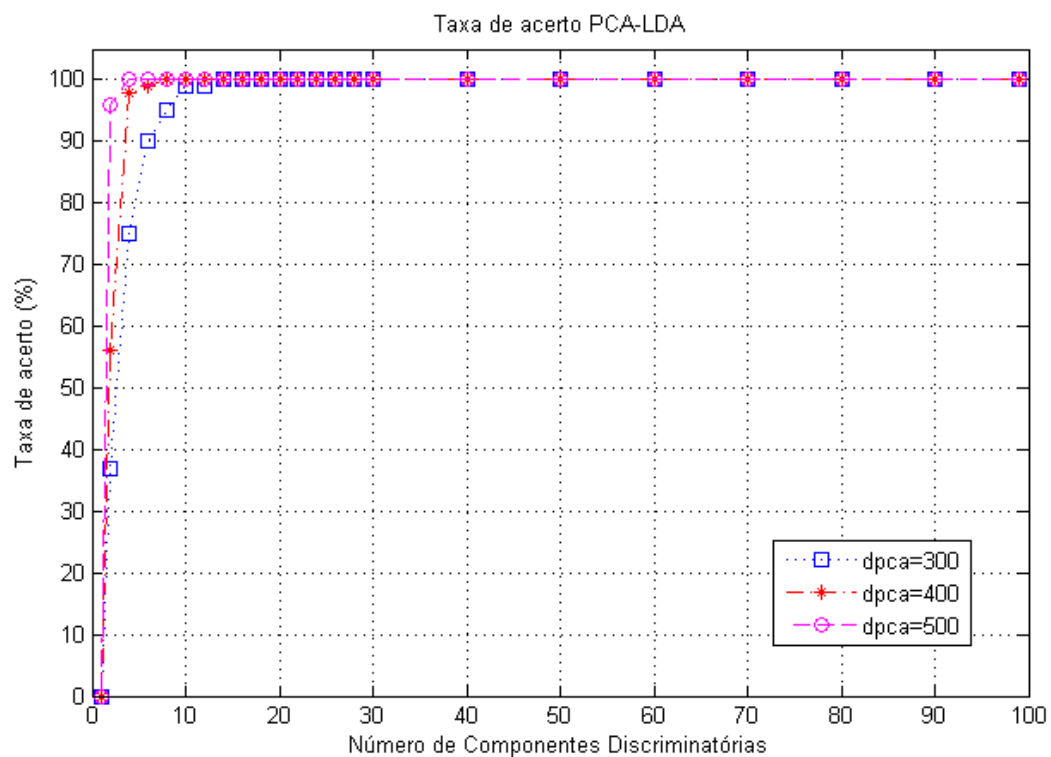


Figura 5.10: Gráfico da taxa de acerto pelo método PCA-LDA para dpca igual a 300, 400 e 500, base de dados CASIA B.

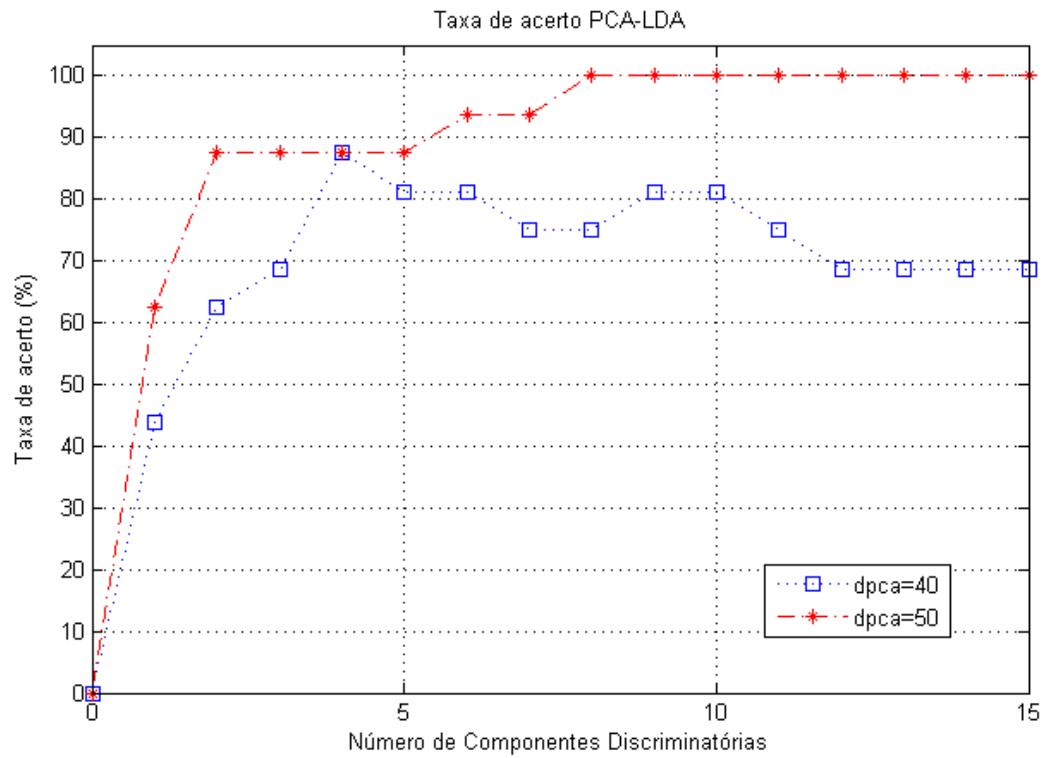


Figura 5.11: Gráfico da taxa de acerto pelo método PCA-LDA para dpca igual a 40 e 50, base de dados GATECH.

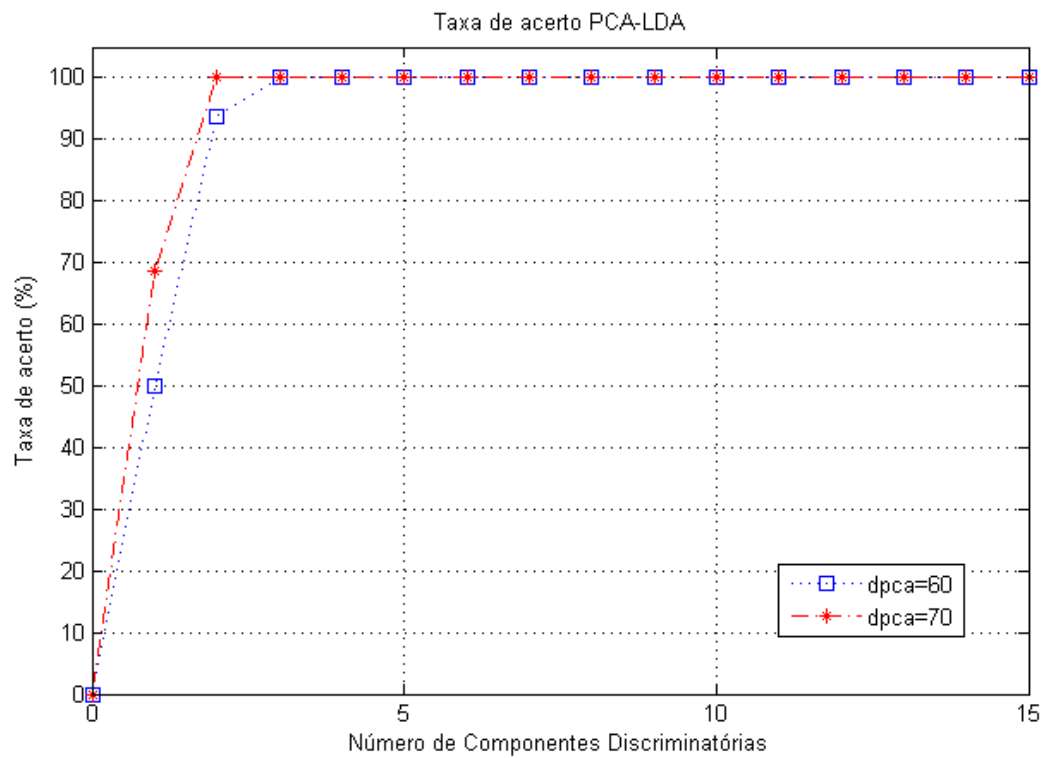


Figura 5.12: Gráfico da taxa de acerto pelo método PCA-LDA para dpca igual a 60 e 70, base de dados GATECH.

Nas Figuras 5.7 e 5.8, vêem-se os resultados de classificação da base de dados CASIA A. Para dpca igual a 100, a taxa de acerto começa a ser satisfatória para 6 componentes, havendo uma tendência central, e partir de 22 componentes a taxa de acerto do sistema começa a decrescer, atingindo seu mínimo em 99 componentes. Para o intervalo que vai de 6 a 60 componentes, tem-se uma taxa de acerto acima de 80%. A maior taxa é de 91% de acerto na classificação utilizando-se 22 componentes. Para dpca igual a 150 e 200, há uma tendência de decaimento menor na taxa de acerto, tendendo a uma estabilização quanto maior for o número de componentes. Para dpca igual a 150, a maior taxa de acerto ocorre em 40 componentes, e esta taxa decresce chegando a 87% para 99 componentes.

Já para dpca igual a 200, a taxa de acerto atinge seu máximo de 100% de acerto chegando a uma saturação, e permanecendo constante para qualquer número de componentes principais acima de 18. Para dpca acima de 200 (Figura 5.8), o número de componentes necessários é bem pequeno, porém existe a desvantagem de se ter dpca muito grande. Todos os resultados alcançam 100% de acerto. Foi testado dpca maior que 500, mas os resultados foram idênticos ao do dpca igual a 500.

Os resultados nas Figuras 5.9 e 5.10 para a base de dados CASIA B foram semelhantes aos anteriores. No entanto, por ser uma base de dados com os indivíduos usando roupas e acessórios diferenciados, o classificador obteve mais dificuldade na classificação, já que a dimensão da PCA e o número de componentes necessários foram maiores, e as taxas de acerto menores que a da base CASIA A. Para dpca igual a 100, a taxa de acerto cresce atingindo seu máximo de 75% em 30 componentes, e decai atingindo 25% com todas as componentes. Para dpca igual a 150 seu máximo é de 92% de acerto com 40 componentes, e decresce até atingir 66% de acerto usando todas as componentes.

Considerando ainda a base CASIA B, para dpca igual a 200 ocorre uma tendência de estabilização a partir de 30 componentes, onde se atinge 100% de acerto nas classificações. Este número de componentes é superior ao da base CASIA A (o que mostra a necessidade que o classificador HMM teve de ter maior número de dados para atingir taxas de acerto maiores). Na Figura 5.10, observa-se que o número de componentes necessários para atingir taxas de acerto iguais aos da base CASIA A, da Figura 5.8, é sensivelmente maior. Por exemplo, para dpca igual a 300, na Figura 5.8 foi necessário no mínimo 8 componentes para 100% de acerto e na Figura 5.10 foi necessário no mínimo 14 componentes.

Nas Figuras 5.11 e 5.12, estão os resultados da base de dados GATECH. Seu resultado foi semelhante aos apresentados da base CASIA A e B, havendo uma tendência de aumento e redução da taxa de acerto para dpca pequeno, por exemplo para dpca igual a 40, mas esta tendência começa a ser modificada quando a dpca aumenta, e a taxa de acerto é melhorada com o aumento do número de componentes. Pode-se observar que para dpca igual a 40, o número de acertos máximo da identificação de indivíduos foi de 87.5 e para dpca igual a 50 chegou-se a 100% utilizando 9 componentes. Atingiu-se também 100% de acerto com dpca igual a 60 e 70, mas com apenas 4 e 3 componentes, respectivamente.

A partir das análises feitas, a conclusão que se chega para estes resultados é que quanto maior a dpca, menor é o número de componentes discriminatórias necessários para atingir altas taxas de acertos. Mas se a dpca for muito pequena, menor que 100, mesmo com um número de componentes muito alto, ou mesmo todas, não se alcançam taxas acima de 90%. Esta escolha de ajuste de parâmetro não é aconselhável para ser utilizada num sistema de identificação indivíduos pela dinâmica do caminhar.

5.4 Método de Redução Dimensional Usando POV

Com base nos resultados apresentados para técnicas de redução dimensional utilizando PCA e LDA, pensou-se num método de redução dimensional que não necessitasse de se realizar uma transformação nos dados de entrada do sistema. Uma destas técnicas é a POV, que calcula um vetor de coeficientes r_m , como apresentado no Capítulo 3, equação (3.14), sendo este uma representação da proporção de variâncias inter e intra-classes (Salomão, 2007). A partir dos valores de r_m calculados, estes se tornam limiares a serem empregados em relação ao número de linhas (neste caso não são colunas) que compõem o vetor de características a ser classificado no HMM. Assim, são eliminadas as linhas cujos coeficientes sejam menos discriminatórios, reduzindo-se o conjunto de dados, e resguardando aquelas que melhor representem uma determinada classe de indivíduos.

Como a matriz que compõe as amostras do conjunto de dados para a base CASIA A e B possuem dimensão 4000×600 e para a base GATECH tem-se dimensão 4000×80, nas Tabelas 5.4, 5.5 e 5.6 é possível observar os coeficientes POV em função do número de linhas a serem selecionadas.

Tabela 5.4: Limiar do coeficiente POV e redução do número de linhas dos vetores, para a base de dados CASIA A.

Limiar de r_m	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
Número de linhas	4000	3995	3887	2748	1553	843	326	85	25	5	3

Tabela 5.5: Limiar do coeficiente POV e redução do número de linhas dos vetores, para a base de dados CASIA B.

Limiar de r_m	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
Número de linhas	4000	3989	3973	3207	1643	594	136	32	12	4	3

Tabela 5.6: Limiar do coeficiente POV e redução do número de linhas dos vetores, para a base de dados GATECH.

Limiar de r_m	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
Número de linhas	4000	2964	1882	1124	552	236	85	27	10	4	3

Através das Tabelas 5.4, 5.5 e 5.6, observa-se que o limiar r_m de cada base de dados é muito específico para cada um de seus conjuntos de amostras, não havendo uma regra que possa se aplicar que relacione limiar r_m e dimensão do vetor de características. Por exemplo, para um limiar de r_m até 0,4, a base de dados CASIA A tem menor número de linhas que a base CASIA B, porém acima de 0,4 a base de dados CASIA B é quem passa a ter menor número de linhas.

Os resultados de classificação de indivíduos utilizando o HMM, para o método de redução dimensional POV estão ilustrados a seguir na Figura 5.13.

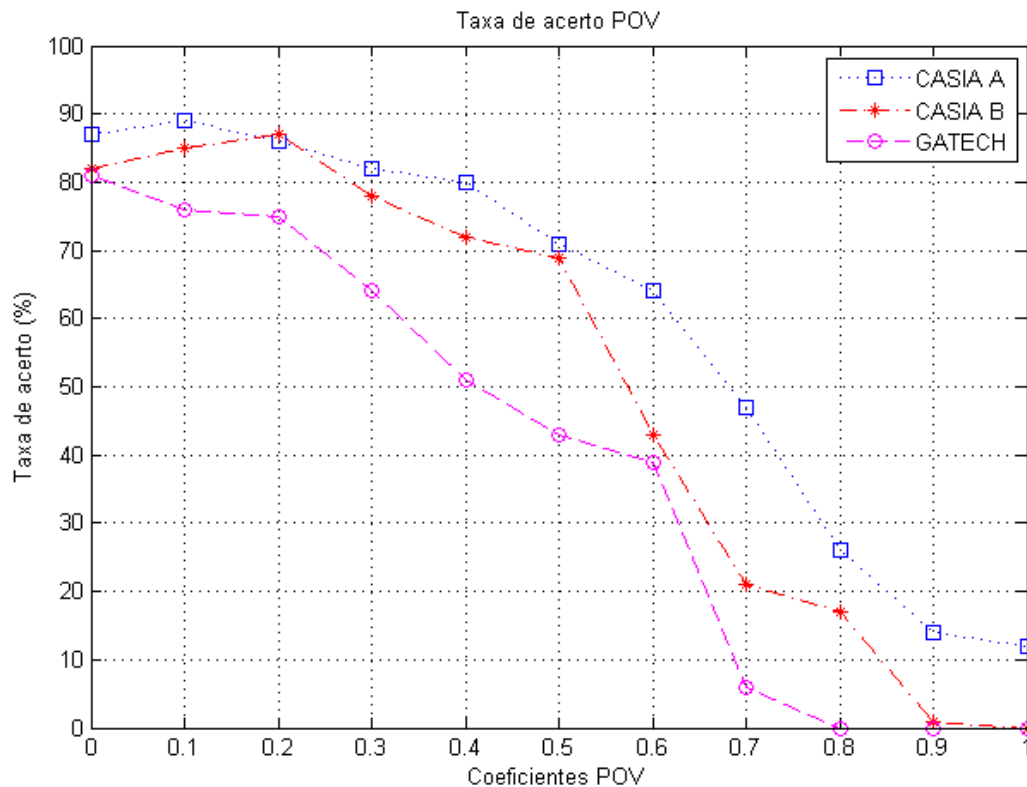


Figura 5.13: Gráfico da taxa de acerto pelo método POV, base de dados GATECH.

Pelo gráfico da Figura 5.13, observa-se que os resultados apresentados utilizando POV como redutor dimensional possuem algumas limitações, pois somente com uma grande quantidade de dados foi possível obter taxas de acertos satisfatórias. Na base de dados GATECH este valor é de 81%, mas fazendo o uso de todo o vetor de características, o que computacionalmente não é atrativo. Já para a base CASIA B este valor é de 87%, utilizando o coeficiente POV de 0,2, sendo que do intervalo de 0 a 0,3, tem-se taxa de acerto acima de 80%. Na base de dados CASIA A, atingiu-se 89% de acerto para um coeficiente POV igual a 0,1, e seu intervalo para taxas de acerto acima de 80% varia de 0 a 0,4. Para valores de coeficientes POV acima de 0,4 a taxa de acerto cai intensamente nas três bases de dados, sendo estes coeficientes valores indesejados de serem ajustados num sistema de reconhecimento humano pela dinâmica do caminhar.

Ainda fazendo um comparativo entre as bases CASIA A e B, a segunda tende a ter valores de coeficientes inferiores à primeira, o que já era de se esperar, mas estes valores são muito próximos para dimensões maiores de dados, mostrando que o classificador HMM tem a

capacidade de reconhecer indivíduos que estejam sujeitos a variações do dia-a-dia (acessórios, bolsas, roupas, etc).

5.5 Considerações Finais

É possível fazer um comparativo entre as técnicas utilizadas de redução dimensional analisando suas taxas de acertos, tamanho do vetor utilizado e especificidades da base de dados.

Na base de dados CASIA A, as informações são de indivíduos que andam normalmente num ambiente controlado, cujas silhuetas foram extraídas com grande precisão. Seus dados de treino e teste possuem muitas informações semelhantes. E com as técnicas de redução dimensionais aplicadas, os resultados foram os melhores, dentre as três bases de dados. Um comportamento semelhante apresentou a base de dados CASIA B, que também é em ambiente controlado, porém nos seus vetores de teste há a presença de indivíduos usando roupas largas e bolsas. Neste caso, o HMM mostrou ser capaz de absorver as informações de dinâmica dos indivíduos, entretanto utilizando uma dimensão um pouco superior. Nessas duas bases de dados, as dimensões nas técnicas PCA e LDA utilizadas é bem inferior ao POV. O POV se mostrou mais aplicável para sistemas de reconhecimento que tenham interesse em reduzir as dimensões dos dados, mas que esta necessidade não seja tão relevante.

Quando o banco de dados é reduzido como o da GATECH, as taxas de classificação ficam sensivelmente comprometidas, pois o número de classes é pequeno. Nas técnicas PCA e POV os resultados não foram tão expressivos, mas o sistema de reconhecimento utilizando LDA se mostrou robusto mesmo com um banco de dados muito limitado, basta observar para os altos valores de acertos obtidos nesta técnica.

O HMM apresentou resultados mais expressivos na técnica LDA. O emprego de PCA-LDA também atingiu bons resultados, pois estas duas técnicas trabalhando em conjunto, conseguem capturar informações representativas de cada classe e informações que as diferenciem das demais classes. Mas vale ressaltar que a técnica PCA-LDA só se justifica sua aplicação nos casos de poucas amostras por classe.

Ainda uma análise interessante a se fazer, é relacionar o tratamento dado as sequências de vídeos para extração das silhuetas dos indivíduos e a utilização destas como vetores de características. Neste trabalho foi utilizada a abordagem holística, ou seja, o uso de informações provenientes da dinâmica do caminhar de indivíduos. Já no trabalho de (Salomão et al., 2006a) foi utilizada uma abordagem holística, porém estática dos dados, onde se fez uso da técnica ESC, já comentada no Capítulo 1. Além disso, foi feita a análise de identificação de indivíduos através da forma de caminhar, da face e estas duas juntas através da fusão dos dados. Seus resultados percentuais de acertos, para o banco de dados GATECH, foram de 89,91% para a forma de caminhar, 93,52% para a face e 90,74% para a fusão de ambas as características biométricas, (inferior ao apresentado neste presente trabalho usando LDA-HMM, que foi de 93,75%). Desta forma, deve-se dar atenção para técnicas que possam aumentar ainda mais as taxas de acerto das características biométricas isoladas, em especial a da forma de caminhar, que obteve uma taxa de acerto superior ao da face.

No trabalho de (Yu et al., 2006), também foi testado o banco de dados CASIA B através da técnica de Energia da Imagem, semelhante ao de (Salomão et al., 2006a), com a finalidade de ser observar como os fatores que caracterizam o banco de dados podem influenciar no desempenho de um sistema de reconhecimento. Comparativamente ao trabalho presente, seu resultado foi inferior, obtendo-se uma taxa de apenas 32,7% para o caso do indivíduo estar usando roupas largas (casaco) e de 52% para o caso do indivíduo estar levando consigo uma bolsa.

Testes realizados em (Kale et al., 2004), utilizando a abordagem holística baseada no contorno da silhueta e empregando o HMM para identificação de indivíduos, apresentaram resultados satisfatórios, de até 89% para a forma de caminhar. Vê-se com isso a importância de se utilizar uma abordagem que possa explicar ao máximo um conjunto de dados, e a escolha de um classificador que seja capaz de captar as informações necessárias para que um sistema de identificação de pessoas seja o mais robusto possível.

Capítulo 6

Conclusões e Projetos Futuros

Este trabalho teve como proposta desenvolver um sistema de identificação de indivíduos através da característica biométrica do caminhar, característica esta distinta entre as pessoas. O principal problema envolvido foi de avaliar o quanto este parâmetro associado ao caminhar humano pode influenciar um sistema automático de identificação e se a abordagem proposta no trabalho é boa o suficiente para dar bons resultados.

Foram mostradas as etapas de pré-processamento envolvidas e necessárias para uma boa extração de dados. A utilização de métodos para estimação de fundo seja em ambientes controlados seja em ambientes não controlados, afetados pela iluminação local, e alguns tipos de filtragens se mostram necessárias para obtenção das silhuetas e para que não haja perda de informações relevantes das silhuetas de cada indivíduo da base de dados (Filtragem Mediana, Filtragem Morfológica e Filtragem Canny). O uso da técnica de compensação de brilho mostrou resultados satisfatórios para aplicações com iluminação controlada. Para aplicações *outdoor* (fora de um laboratório) acredita-se que seu uso seja determinante.

O uso das técnicas de redução dimensional teve a finalidade de mostrar mais claramente como é o comportamento do HMM, principalmente frente às variadas transformações aplicadas aos dados das bases de forma do caminhar utilizadas. Mas os resultados também estão relacionados à etapa de pré-processamento no que tange ao ambiente em que foram realizadas as filmagens e a qualidade das imagens capturadas, e em como estão os indivíduos sobre os aspectos de posicionamento em relação à câmera e aos acessórios ou roupas utilizadas.

Como os vetores extraídos das sequências de imagens possuem alta dimensão, utilizou-se as técnicas PCA, LDA e POV com o objetivo de fazer esta redução de dimensão. Neste sentido, buscou-se avaliar técnicas que ora melhor representam cada indivíduo ora que mais o discriminam em relação aos outros indivíduos que compõem as bases de dados. O classificador estocástico (HMM) utilizado obteve taxas de acerto satisfatórias para alguns valores dimensionais, podendo ser para valores dimensionais intermediários se a aplicação se

justificar (POV) ou mesmo para pequenos valores como os encontrados nas técnicas de PCA e LDA. Os dados quando aplicados a LDA são transformados de forma a salientar as informações mais representativas de cada indivíduo e discriminatórias entre indivíduos.

As taxas mais expressivas estão relacionadas ao método de LDA, mostrando que a dinâmica da forma de caminhar é uma característica distinta em cada pessoa, e que esta característica pode ser uma boa forma de representar cada indivíduo ou mesmo de distinguir indivíduos. Em relação às bases de dados utilizadas, obteve-se também bons resultados para a base GATECH, de pequena população, quanto nas bases CASIA, que são um pouco maiores. O HMM também se mostrou capaz de identificar indivíduos que estejam sujeitos as variações de vestuário e acessórios utilizados, justificando o emprego da dinâmica.

Os resultados obtidos neste trabalho e em (Salomão et al., 2005), (Salomão et al., 2006a) e (Salomão et al., 2006b) mostram que a forma de caminhar tem potencial para se tornar um importante fator ou característica nos futuros sistemas de identificação de indivíduos, podendo esta ser utilizada isolada (somente a forma de caminhar para reconhecimento) ou de forma combinada com outras características biométricas (através de fusão hierárquica). O objetivo é que se consiga obter altas taxas de acertos no reconhecimento de indivíduos para a forma de caminhar isoladamente, de forma que quando esta for utilizada com fusão os resultados sejam ainda mais superiores.

Além dos resultados obtidos através de testes de classificação, vale ressaltar a simplicidade do sistema desenvolvido, quando comparado a outros sistemas de reconhecimento de pessoas. É um sistema de baixo custo e acessível, sendo necessário apenas a aquisição de uma câmera e um computador pessoal. Outra vantagem está no fato de não se exigir a colaboração direta do indivíduo nem mesmo seu contato físico direto com o sensor, como ocorre nos sistemas biométricos em geral. Estes aspectos são muito importantes, pois torna o projeto aplicável e viável economicamente, podendo futuramente atender a sistemas de segurança pública e privada.

6.1 Contribuições da Pesquisa

Esta pesquisa tem como contribuições consolidar a forma de caminhar como uma característica biométrica, que atualmente tem começado a despertar a comunidade científica

de reconhecimento de padrões e visão computacional, e ilustrar aspectos positivos que podem ser aproveitados em sistemas de identificação de indivíduos. Ela revelou ser um parâmetro de suma importância, pois dá um grau de liberdade ao sistema, mostrando não ser tão restrita à distância da pessoa em relação ao sensor de captura e à qualidade da imagem, já que o pré-processamento tem a função de atenuar problemas existentes nas imagens.

Outra contribuição foi de avaliar, em primeiro momento, se esta característica pode contribuir para aprimorar os sistemas de reconhecimento, e os resultados apontam positivamente para esta direção. Como os resultados foram satisfatórios na utilização da forma de caminhar isoladamente, tudo indica que ela também pode ser aproveitada em sistemas que utilizem mais de uma característica biométrica (fusão de características), aumentando as taxas de acerto de identificação, ou mesmo sendo utilizada como um parâmetro discriminatório de reconhecimento.

O trabalho mostra que as técnicas tradicionais usadas para reconhecimento de pessoas pela face podem também ser aplicadas na forma de caminhar, obtendo-se também alto desempenho quando comparadas com outras características biométricas já consolidadas.

O HMM atuando como classificador estocástico se mostrou promissor em amostras de conjuntos populacionais grandes e pequenos, e para variados tamanhos dimensionais.

Do mesmo modo há a contribuição como mais uma referência para trabalhos e pesquisas que envolvam a forma de caminhar, pois poucas são as referências bibliográficas sobre este assunto específico.

6.2 Sugestões para Trabalhos Futuros

Algumas sugestões para trabalhos futuros visando aprimorar o sistema de identificação proposto estão listados abaixo:

- Utilizar bases de dados com número de pessoas maior, para se ter uma aproximação ao se teria na prática num sistema de reconhecimento;
- Criar uma base de dados própria, fazendo capturas em ambientes distintos (internos e externos), em dias distintos, com roupas distintas, e com um

número de pessoas acima de 100, com a finalidade de impor que o sistema a ser testado sobre estes aspectos seja o mais robusto possível;

- Buscar técnicas de pré-processamento que sejam computacionalmente mais rápidas, principalmente no que diz respeito ao processo de extração de fundo, embora a técnica utilizada aqui tenha se mostrado eficiente e robusta na extração do fundo fixo;
- obter as curvas de falsa aceitação / falsa rejeição para cada experimento;
- Utilizar técnicas de seleção e extração de características como a Análise de Componentes Independentes (*Independent Component Analysis* - ICA) e a Transformada Wavelets (*Wavelets Transforms* – WT);
- Buscar outros tipos de classificadores que possam ser relacionados à dinâmica do caminhar, de modo a enfatizar esta característica, como o *K-Means Segmental*;
- Usar treinamento discriminativo para HMM's.

Referências Bibliográficas

- Baum, L. E. (1972). An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic of Markov Process, *Inequalities*, Vol. 3, pp. 1-8.
- Belhumeur, P. N., Hespanha, J. P., Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, *IEEE Transactions Pattern Analysis and Machine Intelligence*, Vol. 19, no. 7, pp. 711-720.
- Bellman, R. (1957). *Dynamic Programming*. Ed. Princeton University Press.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Ed. Springer.
- Bobick, A. F. e Davis, J. W. (2001). The Recognition of Human Movement Using Temporal Templates, *IEEE Transactions Pattern Analysis and Machine Intelligence*, Vol. 23, no. 3, pp. 257-267.
- Canny, J. A. (1986). Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, no. 6, pp. 679-698.
- CASIA, Casia Gait Database B of the Institute of Automation, Chinese Academy of Sciences (Instituto de Automação da Academia Chinesa de Ciências). Disponível em: <<http://www.cbsr.ia.ac.cn>>. Acesso em 29 de julho de 2009.
- Cheng, M., Ho, M. e Huang, C. (2007). Gait Analysis for Human Identification Through Manifold Learning and HMM, *IEEE Workshop on Motion and Video Computing*, pp. 969-972.
- Cunado, D., Nixon, M. e Carter, J. (1997). Using Gait as a Biometric, Via Phase-weighted Magnitude Spectra. *Proceedings of the 1st International Conference Audio-and-Video based Biometric Person Authentication*, pp. 95-102.
- Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*. Boston: Academic Press. 2nd. Edition.

- Ekenel, H.K. ; Sankur, B. (2004), Feature selection in the independent component subspace for face recognition, *Pattern Recognition Letters*, 25, 1377–1388.
- GATECH, Gait database of the Georgia Institute of Technology (Faculdade de Computação do Instituto de Tecnologia da Geórgia). Disponível em: [ftp://ftp.cc.gatech.edu/pub/gvu/cpl/walkers/ subjects](ftp://ftp.cc.gatech.edu/pub/gvu/cpl/walkers/subjects). Acesso em 24 de março de 2008.
- Gonzalez, R. C. e Woods, R. E. (2002). *Digital Image Processing*. Ed. Prentice Hall, 2nd. Edition.
- Hayfron-Acquah, J. B., Nixon, M. S. e Carter, J. N. (2002). Human Identification by Spatio-temporal Symmetry. *Proceedings of the 16th International Conference on Pattern Recognition*. Vol. 1, pp. 632-635.
- Han, J. e Bhanu, B. (2006). Individual Recognition Using Gait Energy Image, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, no. 2, pp. 316-322.
- Jain, A. K., Ross, A. e Prabhakar, S. (2004). An Introduction to Biometric Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 14, no. 1, pp. 1-29.
- Johnson, A., e Bobick, A. (2001). A Multi-view Method for Gait Recognition Using Static Body Parameters, *International Conference Audio-and-Video based Biometric Person Authentication*, pp. 301-311.
- Kale, A., A., Rajagopalan, A. N., Cuntoor,, Kruger, V. (2002). Gait-based Recognition of Humans Using Continuous HMMs. *Proceedings of the 15th IEE International Conference on Automatic Face and Gesture Recognition*, pp. 336-341.
- Kale, A., Sundaresan, A., Rajagopalan, A. N., Cuntoor, N. P., Roy-Chowdhury, A. K., Kruger, V. e Chellappa, R. (2004). Identification of Humans Using Gait, *IEEE Transactions on Image Processing*, Vol. 13, no. 9, pp. 1163-1173.
- Kitane, E. C. e Thomaz, C. E. (2007). Análise de Discriminantes Lineares para Modelagem e Reconstrução de Imagens de Face. *VI Encontro de Nacional de Inteligência Artificial – ENIA*, pp 962-971, Rio de Janeiro, Brasil.

- Kuno, Y., Watanabe, T., Shimosakoda, Y. e Nakagawa, S. (1996). Automated Detection of Human for Visual Surveillance System. *Proceedings of the 13th International Conference on Pattern Recognition*, Vol. 3, pp. 865–869.
- Little, J. J. e Boyd, J. E. (1998). Recognizing People by Their Gait: The Shape of Motion. *Videre: Journal of Computer Vision Research*, Vol. 1, no. 2, pp 2-32.
- Liu, Z. e Sarkar, S. (2005). Effect of Silhouette Quality on Hard Problems in Gait Recognition. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, Vol. 35, no. 2, pp. 170-183.
- Liu, Z. e Sarkar, S. (2006). Improved Gait Recognition by Gait Dynamics Normalization, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, no. 6, pp. 863-875.
- Liu, J. e Zheng, N. (2007). Gait History Image: A Novel Temporal Template for Gait Recognition. *IEEE International Conference on Multimedia and Expo*, pp. 663-666.
- Marques, J. S. (1999). *Reconhecimento de Padrões: Métodos Estatísticos e Neurais*. Ed. Instituto Superior Técnico Press.
- Martínez, A. M. e Kak, A. C. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, no. 2, pp. 228-233.
- Nikolaos, V. B., Hatzinakos, D. e Plataniotis, K. N. (2005). Gait Recognition: A Challenging Signal Processing Technology for Biometric Identification, *IEEE Signal Processing Magazine*, Vol. 22, no. 6, pp. 078-090.
- Phillips, P. J., Sarkar, S., Robledo, I., Grother, P. e Bowyer, K. (2002). The Gait Identification Challenge Problem: Data Sets and Baseline Algorithm, *Proceedings of the 16th International Conference on Pattern Recognition*, Vol. 1, pp. 385-388.
- Rabiner, L. R. (1993). *Fundamentals of Speech Recognition*. Ed. Prentice Hall.
- Vieira, R. M. T., Salles, E. O. T., Salomão, J. M. (2009). Identificação de Indivíduos pela Dinâmica do Caminhar. *IX Simpósio Brasileiro de Automação Inteligente – IX SBAI*, Brasília, Brasil.

- Salomão, J. M., Almeida, A. R. e Salles, E. O. T. (2005). Avaliação de Desempenho de um Classificador SVM no Reconhecimento Pessoal Via Forma de Caminhar com Seleção de Características Através das Transformadas Wavelets. *VII Simpósio Brasileiro de Automação Inteligente – VII SBAI*, São Luis, Brasil.
- Salomão, J. M., Almeida, A. R. e Salles, E. O. T. (2006a). Reconhecimento Pessoal Através da Fusão Hierárquica das Características da Forma de Caminhar e de Imagens Faciais. *XVI Congresso Brasileiro de Automática - CBA*, Salvador, Brasil.
- Salomão, J. M., Almeida, A. R. e Salles, E. O. T. (2006b). Uso das Características Biomecânicas do Caminhar na Identificação Pessoal Através de Redes RBF. *XX Congresso Brasileiro de Engenharia Biomédica*, São Pedro, Brasil.
- Salomão, J. M. (2007). Implementação de Métodos Biométricos Bi-modais Baseados em Fusão de Características para Reconhecimento de Indivíduos. Tese de Doutorado, Programa de Pós-Graduação em Engenharia Elétrica, UFES.
- Souza, J. A. (1999). Reconhecimento de Padrões Usando Indexação Recursiva. Tese de Doutorado, Programa de Pós-Graduação em Engenharia de Produção, UFSC.
- Stewart, C. V. (1999). Robust Parameter Estimation in Computer Vision, *SIAM Review*, Vol. 41, no. 3, pp. 513-537.
- Wang L., Hu, W. e Tan, T. (2002). A New Attempt to Gait-based Identification. *Proceedings of the 16th International Conference on Pattern Recognition*. Vol. 1, pp. 115-118.
- Wang, L. e Tan T. (2003). Silhouette Analysis-based Gait Recognition for Human Identification, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, no. 12, pp. 1505-1518.
- Yam, C.. Y., Nixon, M. S. e Carter, J. N. (2002). On the Relationship of Human Walking and Running: Automatic Person Identification by Gait. *Proceedings of the International Conference on Pattern Recognition*, pp. 287-290.
- Yu S., Tan, D. e Tan, T. (2006). A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition, *XVIII International Conference on Pattern Recognition – ICPR*, Vol. 4, pp. 441-444.

- Zao, W., Chellapa, R. e Krishnaswamy, A. (1998). Discriminant Analysis of Principal Components for Face recognition, *Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, pp. 336-341.
- Zhou, X. e Bhanu, B. (2006). Integrating Face and Gait for Human Recognition, *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, pp. 55.